Quality Control and Improvement with MINITAB
Prof. Indrajit Mukherjee
Shailesh J. Mehta School of Management
Indian Institute of Technology, Bombay

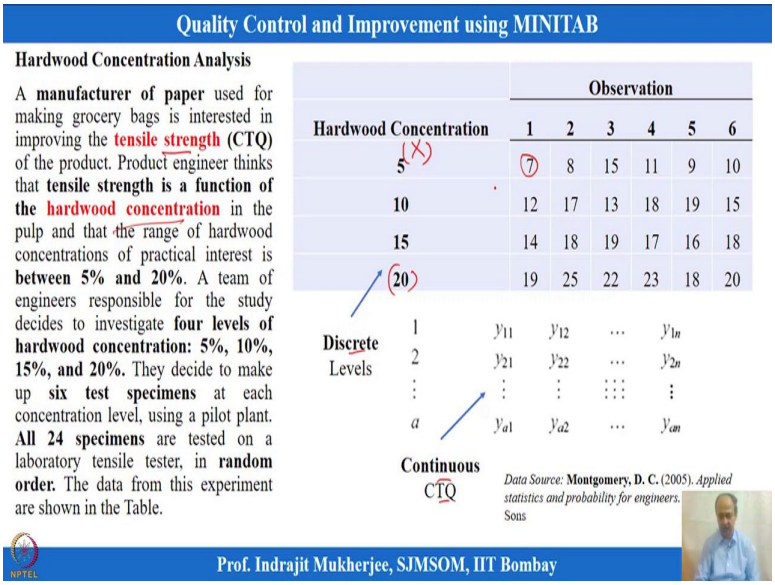
Lecture - 22
One–way ANOVA (Continued)

Hello and welcome to session-22 of our course on Quality Control and Improvement with MINITAB. I am Prof. Indrajit Mukherjee from Shailesh J. Mehta School of Management, IIT Bombay. So, we are discussing about analysis of variance with some examples, we are illustrating how to do it in MINITAB. So, basically we are in interface of control and improvement.

So, this is to identify variables when I am doing analysis of variance, which factor is basically influencing the output or CTQs. We want to identify critical variables like in cause and effect diagram what we have identified, and whether they significantly influence the mean of the CTQ response.

So, intentionally, we will change the condition of X, and we want to see what the influence on the mean of the CTQs or of the CTQsvariance is. So, ANOVA analysis can be used for that.

(Refer Slide Time: 01:08)

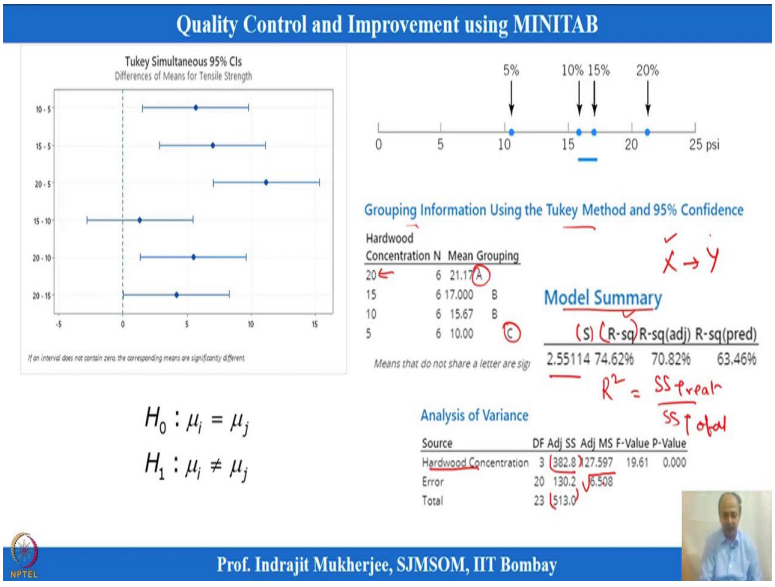


So, we have taken a hardwood concentration example, where say the CTQ is tensile strength. We have seen that how to do the analysis and determine that 20 is the best level of hardwood concentration so that to maximize the CTQ response over here.

So, we have used a combination of ANOVA analysis and we have used a multiple comparison test, Tukey's test to confirm that which level I should freeze. But although these factors we have identified that changing this factor, this is an important X factors which needs to be considered in further experimentation when we consider more number of Xs like that.

And the condition of ANOVA analysis what we have assumed is that these are at discrete levels, and response, which is tensile strength, is continuous variable. Without continuous kind of these assumptions over here, ANOVA analysis cannot be done. So, value of response should be continuous then only we can do ANOVA analysis like that ok.

(Refer Slide Time: 02:04)

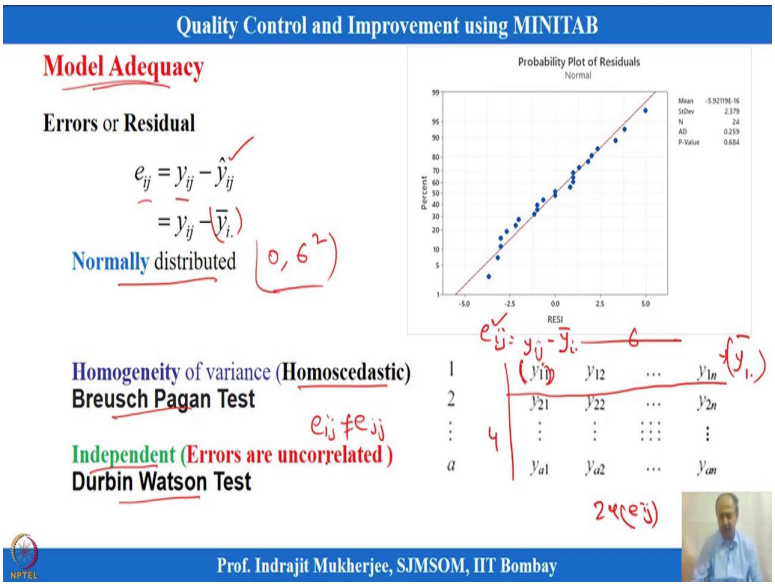


So, for this analysis, we have some information that we have gathered like that grouping information over here for the two case tests, and we have seen the letter code. If they are different; that means, their different levels are having significant mean difference, so that is why we have reached at 20.

And more information what you will get is that this model summaries over here, out of that one summary I will explain over here which is known as R^2 information which is nothing but $\frac{SS_{Treatment}}{SS_{Total}}$.

Why this is important over here? This, explains that how much of the X is variability of Y or the total data set that we have collected over here during experimentation, is explained by this change in X basically. And we are not talking about R^2 adjusted or R^2 predicted that we will talk about when we are talking about regression analysis. And in this S values what you see over here the lower the values, that means, model is quite adequate. This S measures is nothing but square root of this mean square error that you have seen over here. So, square root of this will give you the SS information over here. So, variance of the Y, standard deviation of the Y or residuals we can think of is what we get out of this ANOVA analysis that we are getting over here.

(Refer Slide Time: 03:39)



You have to remember that there are some model assumptions that are needed to be checked over here. After you have completed the analysis, there is a residual analysis that has to be done. What are the residual assumptions? Assumptions are error follows normal distribution with mean zero and standard deviation σ^2. σ^2 can be estimated as mean square error that we are estimating. So, σ estimation can be get out of that. So, normality assumptions of the residual; what is residual over here?

Residual is calculated as $e_{ij} = y_{ij} - \overline{y}_{i.}$. y_{ij} is the response value at i^{th} factor level in j^{th} experimental run. And each of the level average is given by $\overline{y}_{i.}$. So, I will get 24 error terms over here (6 replicates and 4 factor levels). And this can be stored in MINITAB and can be verified whether it is following normal distribution or not.

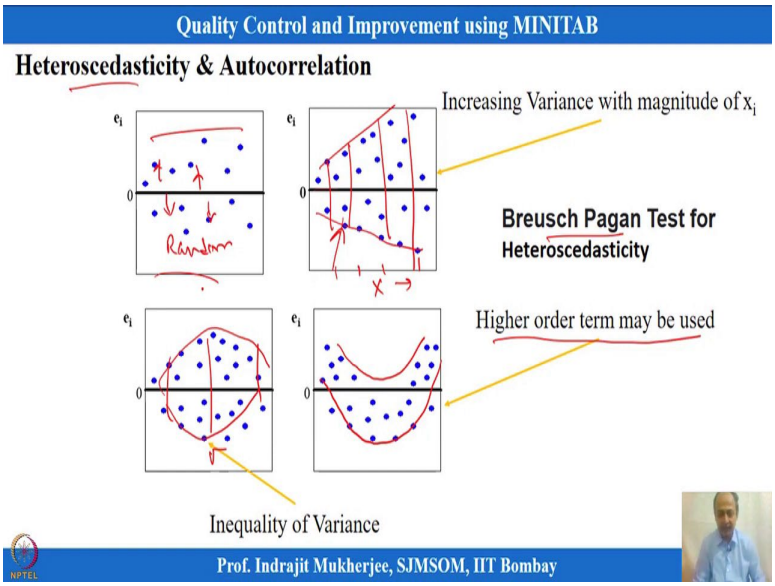
Then the second condition that we can we have to verify is that whether the residuals are homoscedastic or not. So, that means variance of the residual does not change with levels of X. So, variance of the residuals needs to be checked over here. This is one check that we have done initially when we implemented analysis of variance. We have seen the method of checking equal variance and we have used Bartlett test, if the underlying distribution of each group is normal and otherwise we have used Levene's test or multiple comparison test, to confirm whether the variance is same or not. So, similar test exist also which is another test that can be done which is known as Breusch-Pagan test for the residual and which is commonly used by people when we talk about regression.

But we can use the other concepts of Levene's test also to do this homogeneity of variance of the residual or on the y. So, this is Breusch-Pagan test is not available in MINITAB. So, what we have to do that we can do this test in R interface. Durbin-Watson test is available another important assumption is independency of the error. So, error $\left(e_{ij},\ \forall i \neq j\right)$ is independent of any other error.

Auto correlation is another important concept that comes into when we talk about independence of error. We need to check whether the errors are auto correlated or not. So, for that, there is a Durbin-Watson test statistic which is used.

Now, generally error should be independent in when we are doing some test, but in chemical processes sometimes what happens is that errors may be correlated. So, in that case there is a Durbin-Watson test which is possible in MINITAB, but only thing is that value has to be compared with tabulated values. So, that is some time point it is difficult, but there is a process in doing that. So, Durbin-Watson test can be done in MINITAB also. So, in this case, there is a possibility.

(Refer Slide Time: 07:41)



And what do you mean by this heteroscedasticity that we talked about, that is homogeneity of variance checking over here what we are doing by Breusch-Pagan test. And in this case, what we see is that whenever the variance of errors increases in magnitude and this is the error direction and this is the levels of X that we are setting over here.

And if you see that, these are different levels 1, 2, 3, 4, four levels are there let us say and the variance of this is changing what you see over here. So, this is the non-constancy of variance that is observed in the error and this is not recommended. Whenever this exists that means, the interpretation may be wrong and in that case judgment may be also be wrong.

I have rejected the null hypothesis which may not have been rejected like that, so that can be one thing we have to take care of. So, whenever there is a heteroscedasticity condition over here and that is checked by either Breusch-Pagan test or other Levin's test or some other test where non-constancy or variance can be proved, whether existing not existing statistically.

So, in that case, I need to do some correction over there before I make an interpretation of the ANOVA analysis. So, there will be variance transformation, variance stabilizing transformation on the y variables or the CTQs like that. And after that that ANOVA analysis will be done.

So, there is that one of the two options that I have shown you is Box-Cox transformation, another one is Johnson's transformation can be used for variance stabilization. Like, if it is not normal we are using that, so, in case of non constancy of variance also we can use those transformation, and so that the y distribution becomes normal and then we can apply and we can see whether there is the outcome residual is quite noise or not.

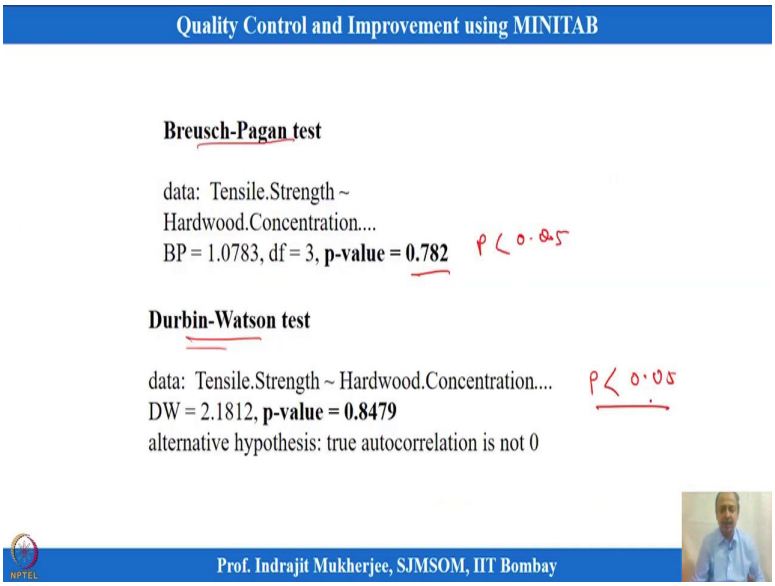
So, that we have to ensure that residual is following normal distribution or not and constancy of variance, and they are independent all these tests. So, Durbin-Watson test statistics that I told can be used for independency test. So, if the behavior is like this, this is basically random.

So, this is expected to be random throughout on the zero-axis or zero on higher side and lower side like this. And if the scenario is like this also variance is changing at different levels of X.

And there can be certain scenarios where errors are like u shapes over here. This indicates that higher order term may have to be introduced. We will discuss about this afterwards. So, in this case maybe non-linearity exists in the models like that which is assumed to be linear that is not the case.

So, in this case nonlinearity, so, inequality of variance is checked over here. We expect a scenario where everything is fine and heteroscedasticity is not there. So, this is the visual impression we can get when you plot the residuals basically.

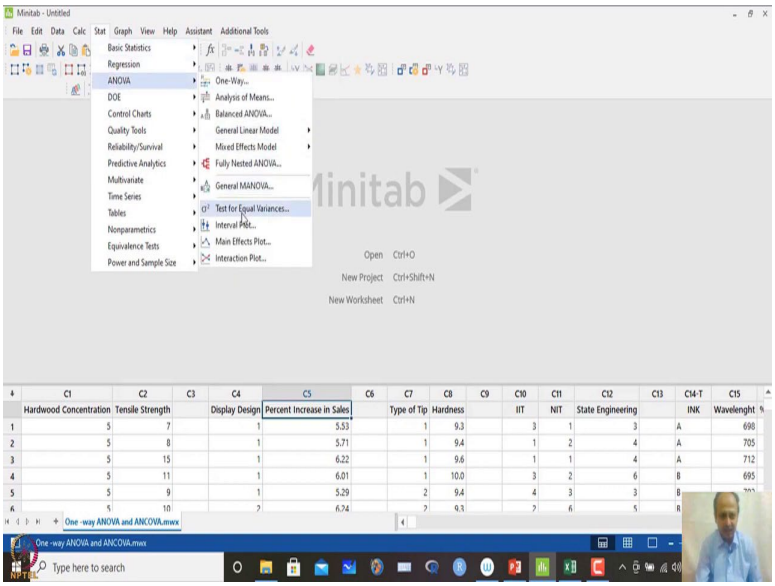
(Refer Slide Time: 10:33)



So, then and that is possible in MINITAB we will get different options to plot the residuals. So, one is I told Durbin-Watson test statistics that is also possible in MINITAB that is also possible in R interface, and Breusch-Pagan test and the p value will indicate if there is significant heteroscedasticity. If p value is less than 0.05, we will say heteroscedasticity is there p value is less than 0.05 over here, we will say autocorrelation exist, the errors are not independent like that.

So, these are the two tests that can be used for interpretation. And let me take one more examples to complete ANOVA analysis, and then we move ahead with some other examples, where this will be violated this, and then we have to see some stabilization options like that or convert into normal distributions like that what are the options like that.

(Refer Slide Time: 11:23)

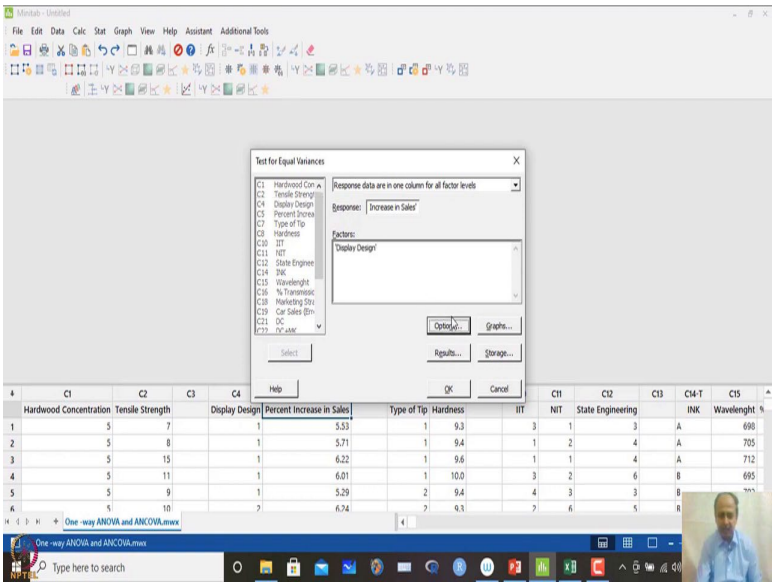


So, let me take another examples to illustrate. So, this is the display design and percentage increase in sales. If you go to a let us say mall or something like pantaloons or somewhere, you are entering and you will say that the structure or design the way you experience. The different products like that will be changed a time and again like that.

So, aisle design can be changed like that. So when they change the design, they expect more sales. So, they keep on updating this monthly, weekly, they change the designs how the people will move in the shops, so that more exposure will happen and more sales will happen like that.

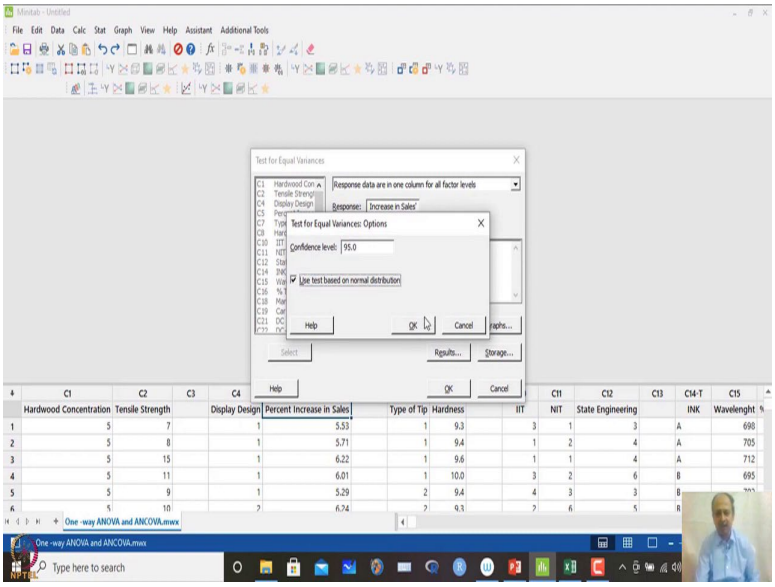
So, one experiment was done with three different display designs over here, and percentage increase in sales which is y or CTQ that we are interested that was monitored over here. So, this is a marketing example we are taking over here, how the display designs are changed to increase the percentage increase in sales, and try to see which design is optimal over here.

(Refer Slide Time: 12:33)



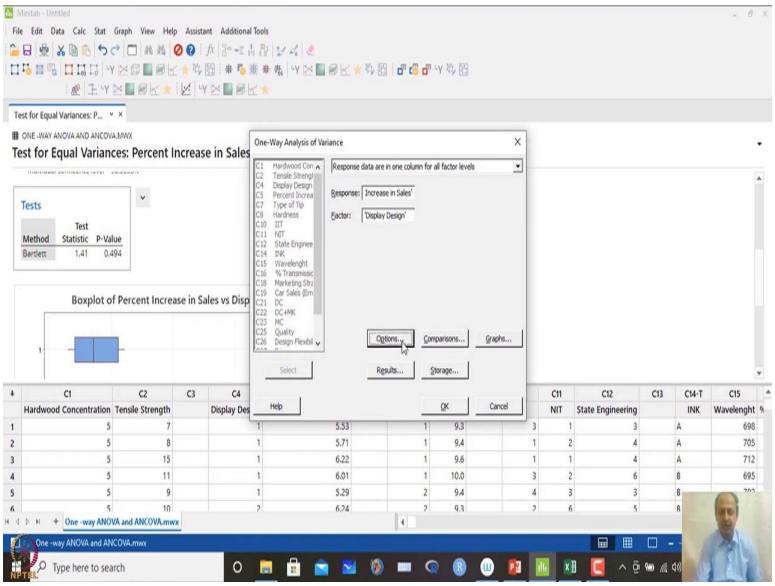
So, in this case what we can do is that, first we can see the variance test to do the basic things. So, ANOVA analysis test of equal variance first test that we are doing over here. And then what we can do is that response are in one column, and then we will say that percentage increase in sales is the one, and this is the factor is display design like this.

(Refer Slide Time: 12:52)

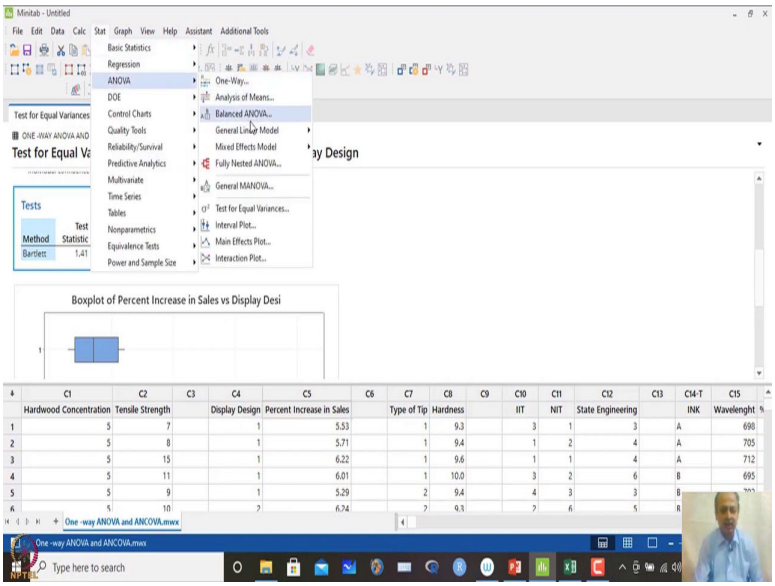


And in options let us assume over here, all the group informations are normally distributed. And we can check this one, so whether they are in groups, whether they are normally distributed or not that check we can do.

(Refer Slide Time: 13:11)



(Refer Slide Time: 13:22)

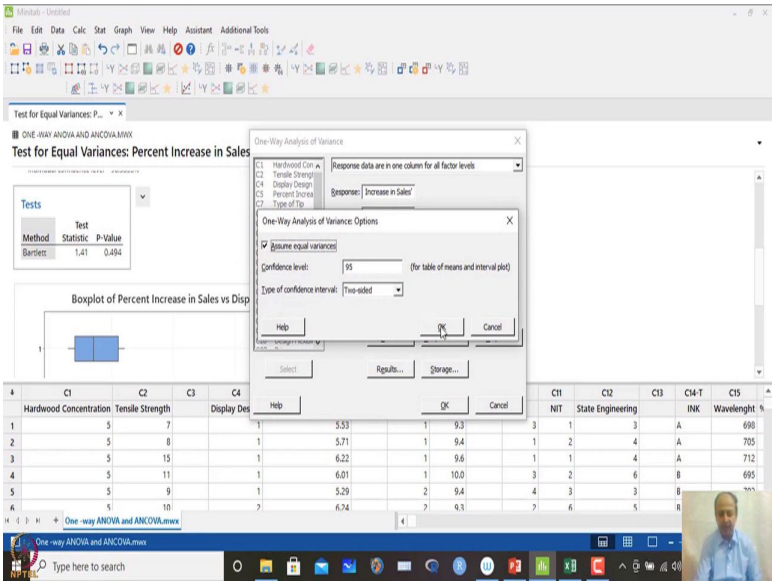


And if you click ok over here what we can do is that, now in this case we are not saving anything. We want to see whether the variance is same and the p value for Bartlett test that I mentioned, is more than 0.05. So that means, the variance is constant.

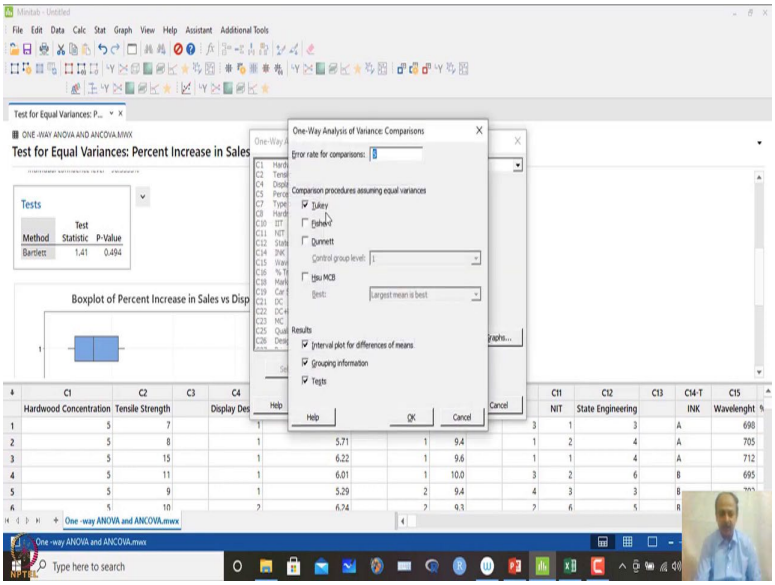
So, there is no problem with that then we go ahead. So, there is no heteroscedasticity that we observe at present. And then we go to one way analysis of variance and then in this case we change this one and response also we are changing over here. And so, the

response is basically percentage increase in sales, and this is the display design over here.

(Refer Slide Time: 13:43)

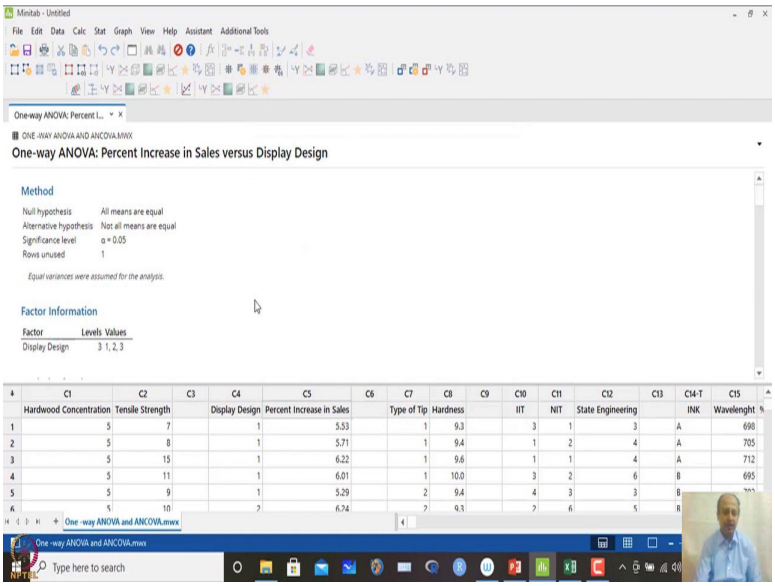


(Refer Slide Time: 13:48)

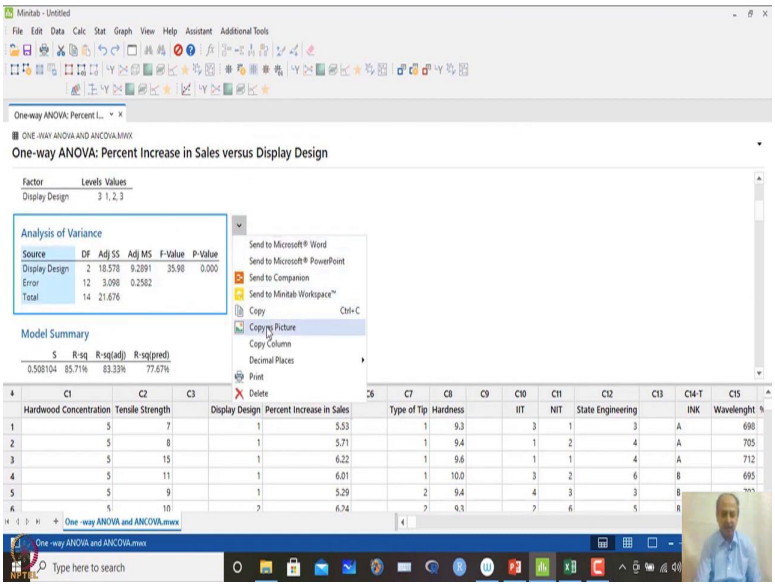


And in options what we have assumed is assume equal variance that is proved already. So, I click ok over here, comparison test two case test I am doing over here. So, and the test information will be given like this. So, if I click ok over here.

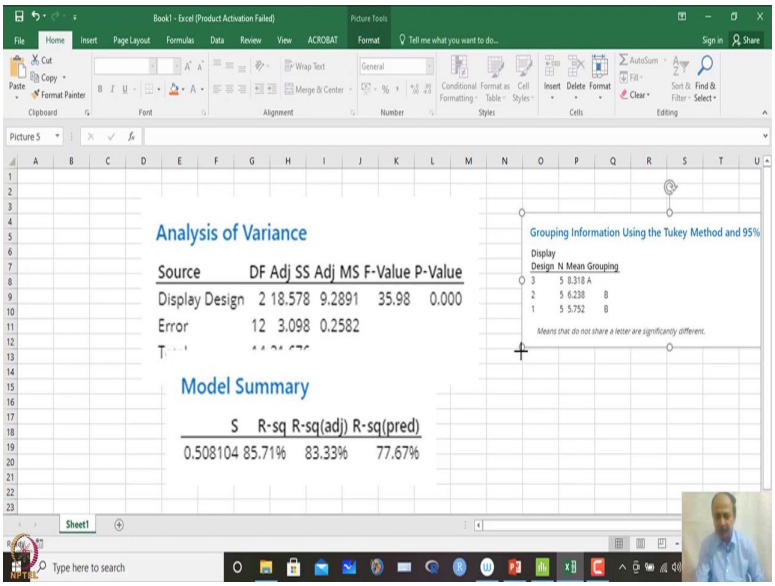
(Refer Slide Time: 13:54)



(Refer Slide Time: 13:55)



(Refer Slide Time: 14:02)

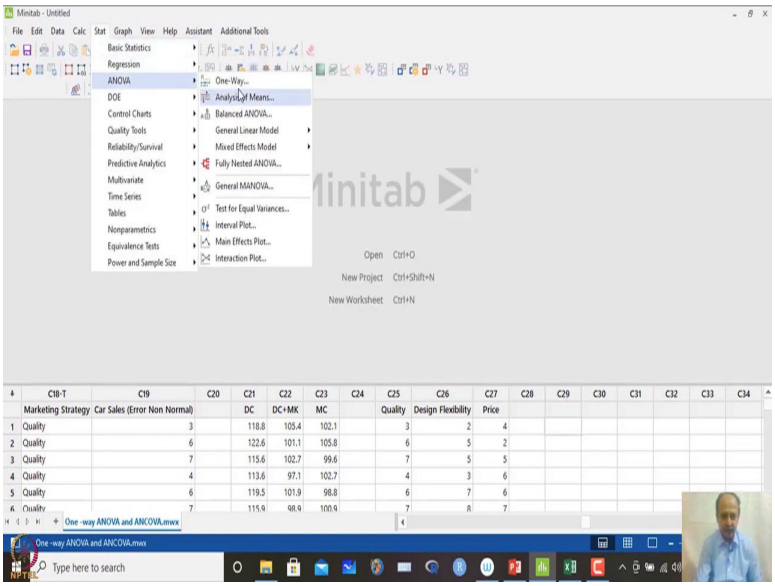


So, what will happen is that I will say whether the design is significant or not. So, I can copy this as a picture over here, and I can just paste it in excel and to enlarge this one and see what is the result. So, display design, the variability of display design is significant or when I change the display design, what is happening is that mean is changing basically at least two levels like that.

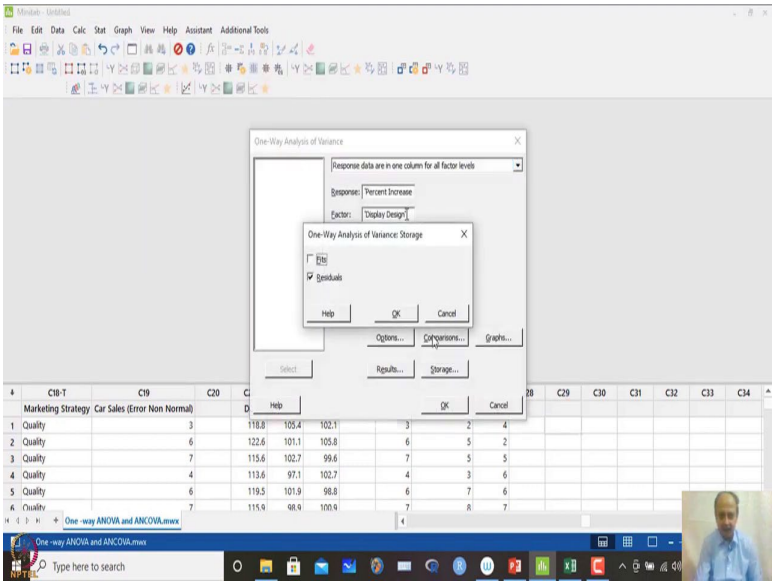
So, this is significant. ANOVA analysis says that display design is an important variable. When you change that one, basically percentage increase in sales also changes. And the R^2 is around 85 percent over here, what do you observe over here.

So, if I paste this one over here information what it says is that basically, R^2 that we are getting over here that 85 percent of the variability of y is explained by this change in X over here, so that is the interpretation that we have from this data analysis. So, in this case, what we can do is that we have also saved the residual. Let me see whether the residual is saved or not.

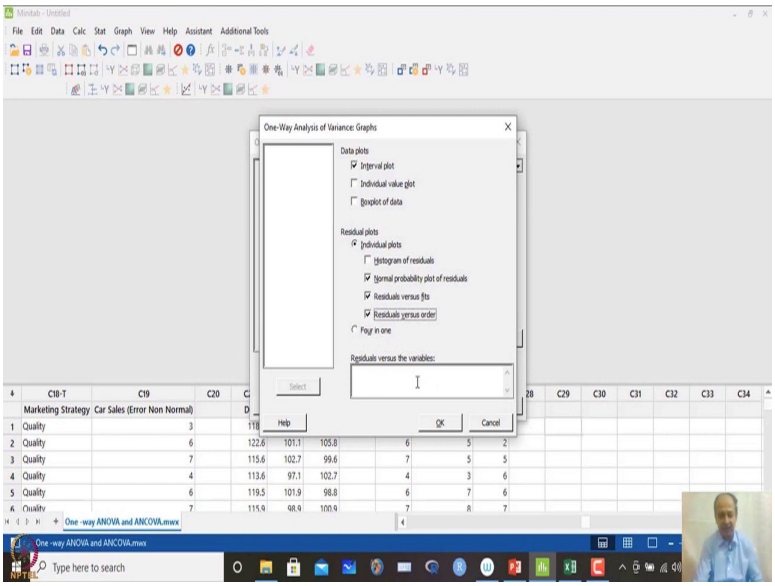
(Refer Slide Time: 15:01)



(Refer Slide Time: 15:05)



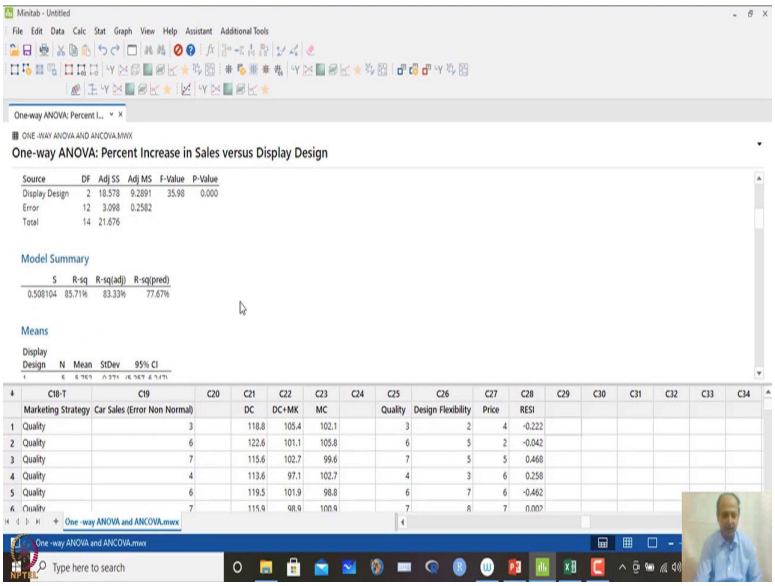
(Refer Slide Time: 15:13)



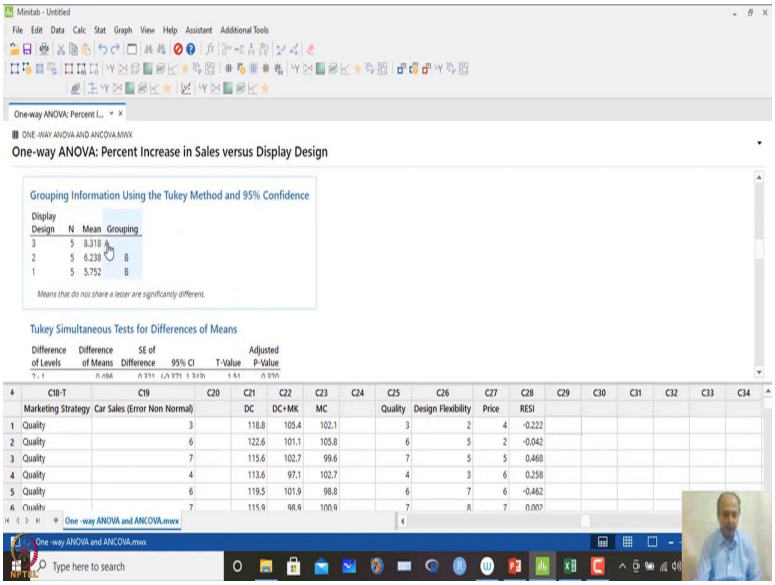
Let us redo this one again, so that I understand at the end the residual will be saved. So, in this case storage, I will just ensure residual is saved over here. So, if you click ok and in the graph what you can do is that I can see normal probability plot of the residual to understand whether the residuals, I am doing the checks and residual versus fit this is one analysis where we can see heteroscedasticity is there or not.

And autocorrelation whether it exists or not the residual versus order this gives you some information visual impacts like that. But Durbin Watson test and Breusch-Pagan test and all these tests can be used to confirm that one statistically, whether there is any difference like whether there is any significant heteroscedasticity or autocorrelation that exists ok.

(Refer Slide Time: 15:47)



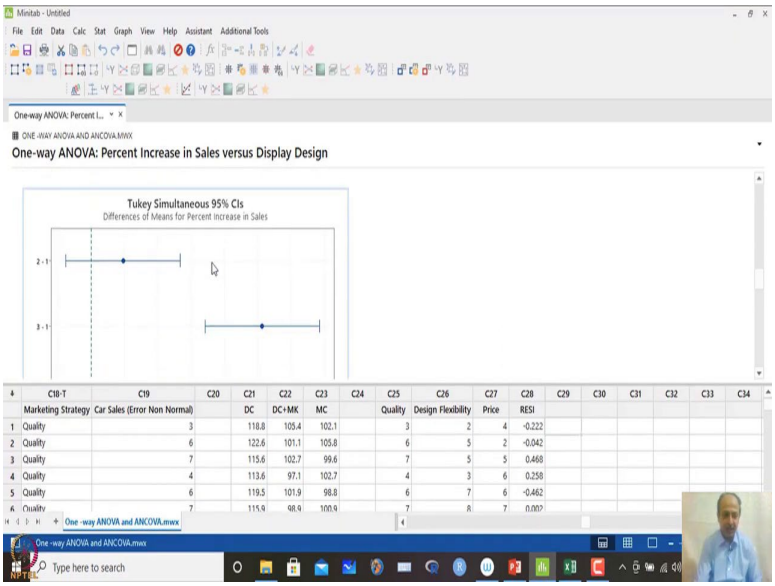
(Refer Slide Time: 15:54)



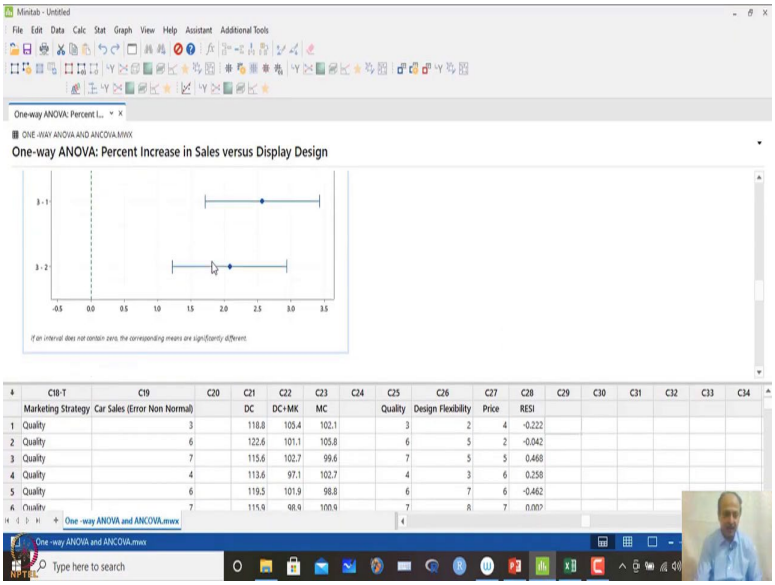
So, in graph, if I do ok over here, so, what we will do is the we can see is that it is significant and then what we can see is that in grouping information that display design 3 is giving a percentage increase in sales which is much higher as compared. Because group code is A and that is different from B codes 2 and 1.

Both 1 and 2 is giving me a lower mean over here. So, if you have to maximize your percentage increase in sales, I have to go by display design 3 over here

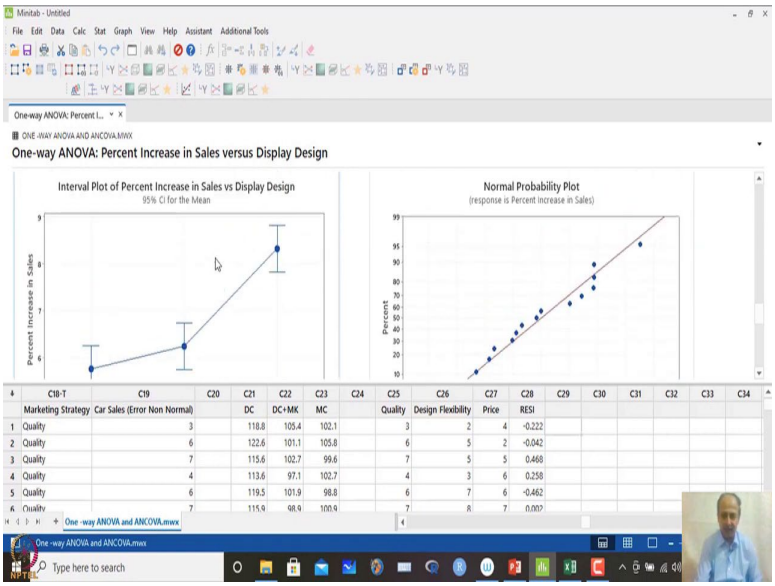
(Refer Slide Time: 16:32)



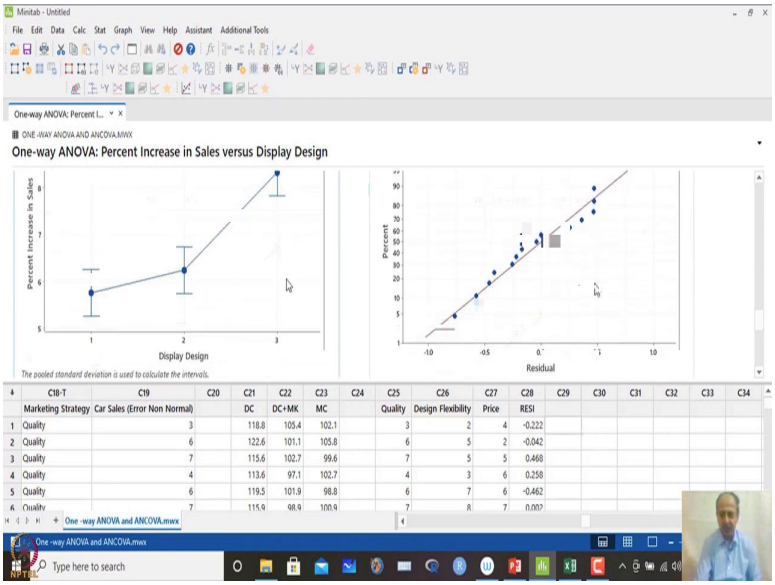
(Refer Slide Time: 16:33)



(Refer Slide Time: 16:34)

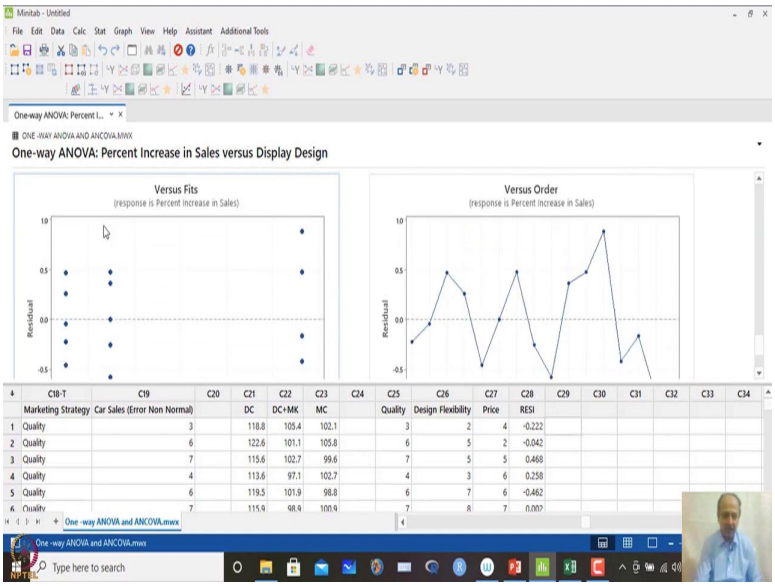


(Refer Slide Time: 16:37)

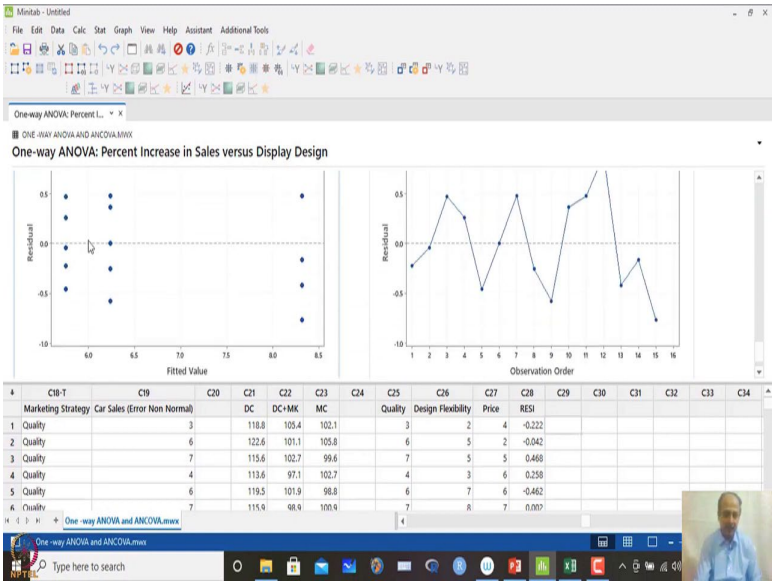


So, in this case, this is possible and then this is two case comparison test and then normal probability test more or less seems to be ok.

(Refer Slide Time: 16:41)

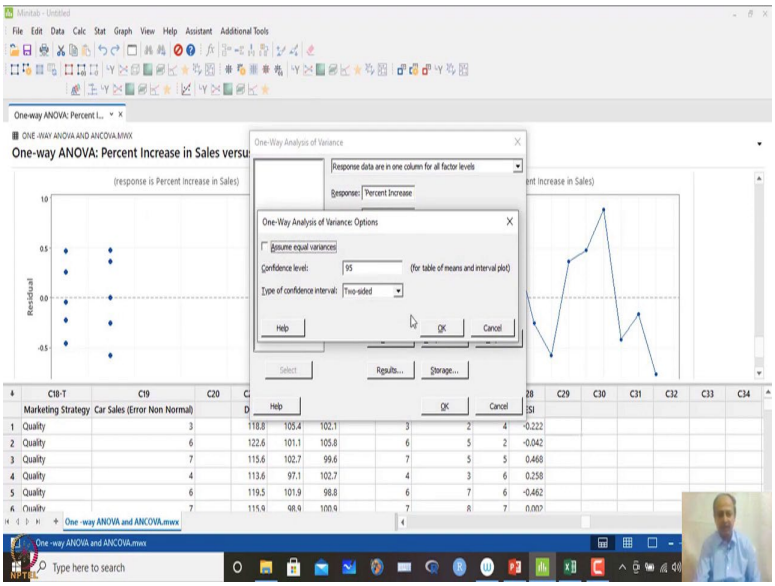


(Refer Slide Time: 16:42)



And then in this case also, this does not seems does not have heteroscedastic behavior, which was also confirmed in earlier case we have done the variance test that is Levene's test. We have done and Bartlett test we have seen also that because, we have assumed normality, so that is why Bartlett test we have done. So, and also there may not be residuals with ordered information also we do not see much changes over here. So, maybe that test also will not fail. So, in this case so maybe no autocorrelation that exists.

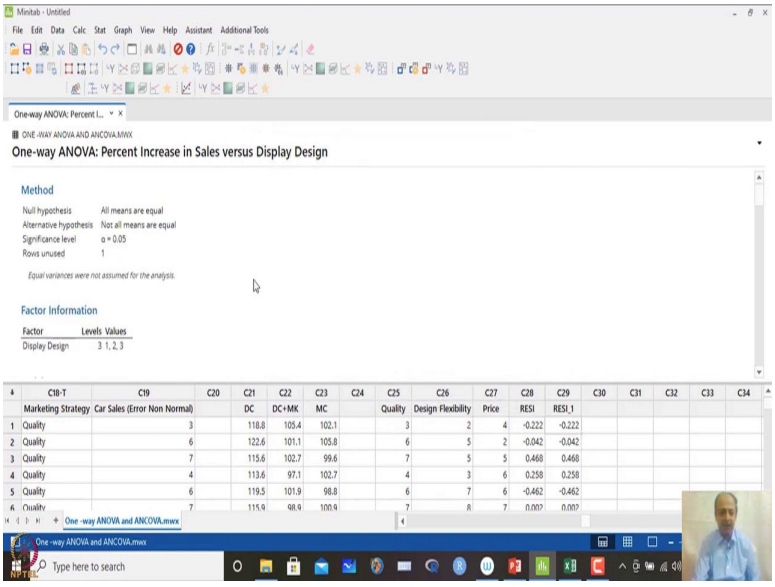
(Refer Slide Time: 17:19)



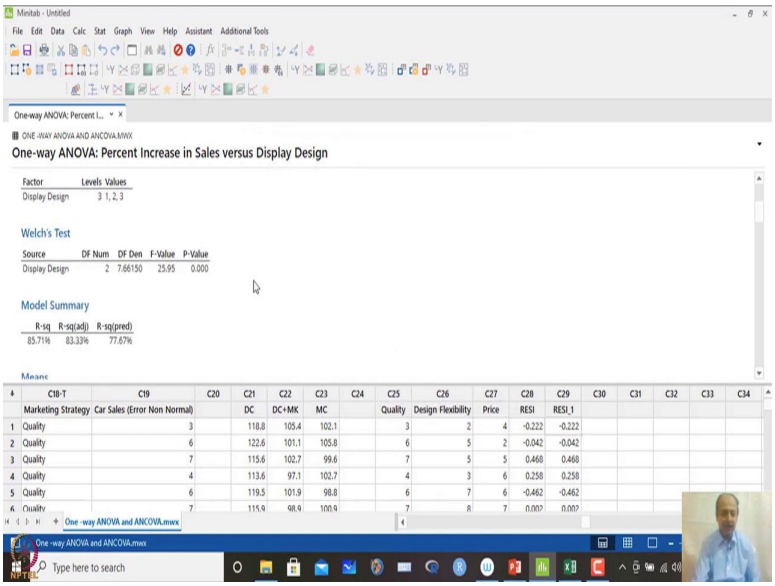
So, in case there is a non-constancy variance. So, in that case, what you have to do one way ANOVA analysis. So, in this option, you have to assume unequal variance over here, so just do not click this option. So, remove this one. And when you do the multiple comparison test, it will ask for this Games-Howell test over here.

So, this you have to click over here, because it only gives Games-Howell test which is the combination I will use with Welch's test whenever the variance heteroscedastic behavior is observed like that.

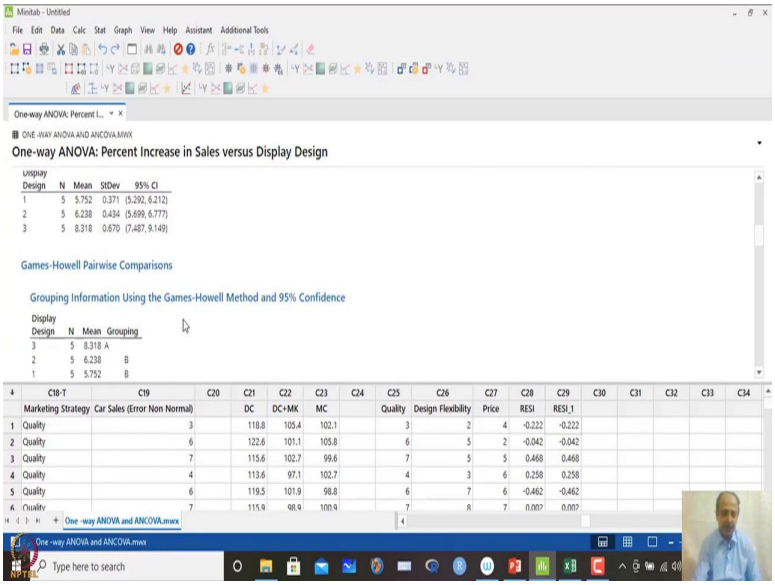
(Refer Slide Time: 17:51)



(Refer Slide Time: 17:51)

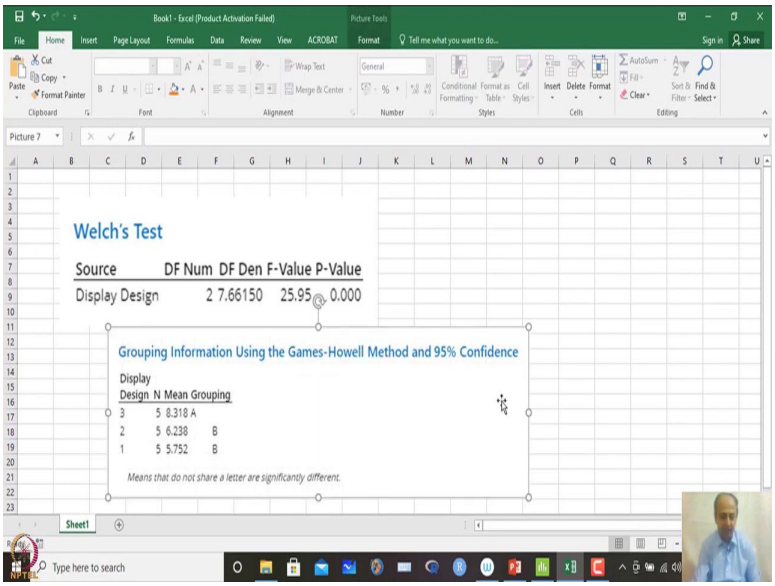


(Refer Slide Time: 17:52)



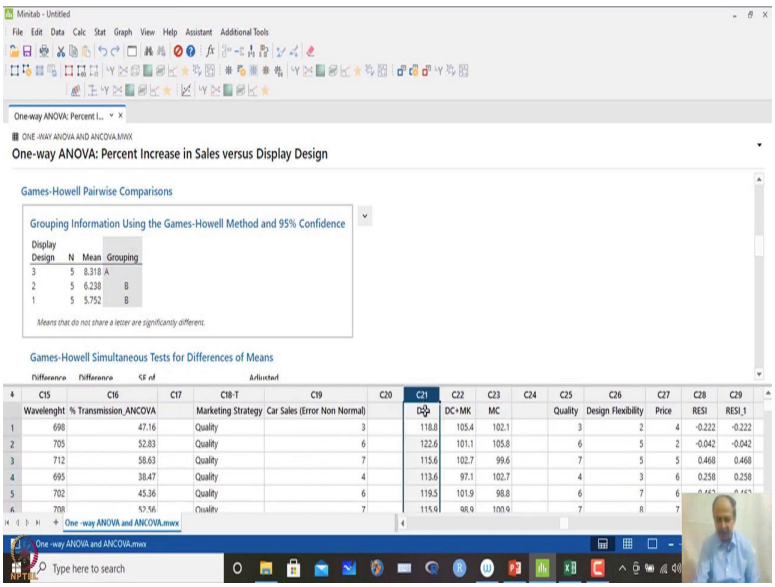
So, if you click ok over here and that takes care of heteroscedastic information. And if you click ok over here, what you will get is that Welch statistics will be given over here. So, you will get this information copy as picture over here. And I can just paste the information.

(Refer Slide Time: 18:04)



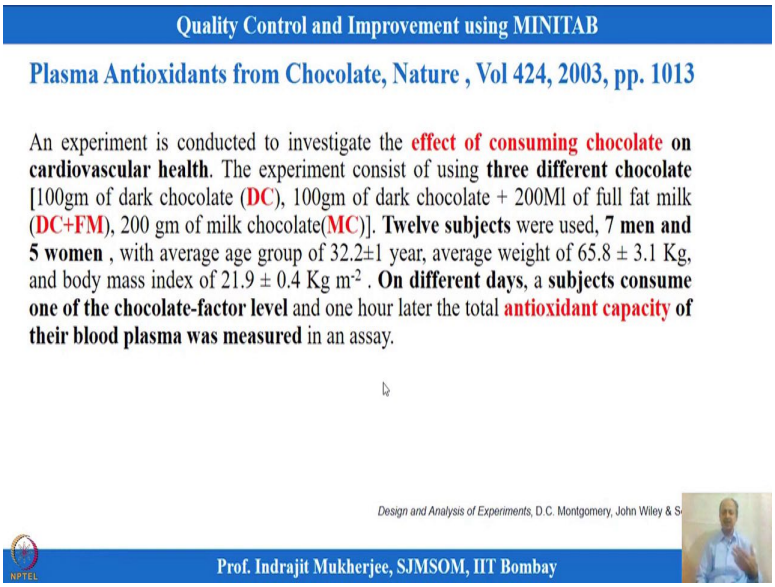
So, what we get from Welch's test in case variance is not constant. So, in this case, display design is significant over here. And it will also give you the information that when I use Games- Howell pairwise comparison. So, here also you will find the information which is different from which one like that. So, I will delete this one, I will delete this one Welch's test in combination of this group information that is Games-Howell method like that. And, sometimes what happens is that you may have to you may be given some information like this.

(Refer Slide Time: 18:58)



For e.g. the data can be saved in columns like this. These are the different types of chocolate, milk chocolate, dark chocolate, and then combination of these. Some CTQ, cholesterol level, which is important over here is measured like that.

(Refer Slide Time: 19:27)



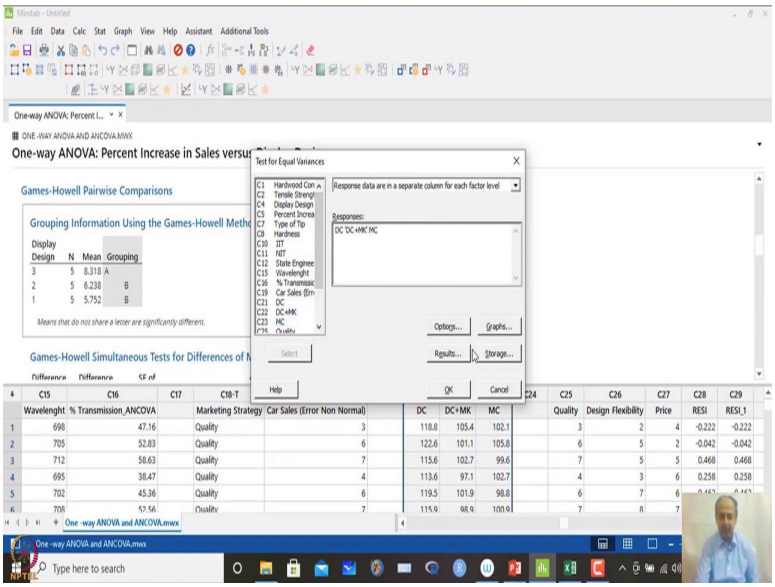
And I want to see whether there is any difference when I consume completely pure dark chocolate, milk chocolate or their combination. So, this example is taken from again from Montgomery's book. So, this is the example that I am talking about over here, experiment is conducted to investigate the effects of consuming chocolates on cardiovascular health.

Over here what is monitored is that antioxidant capacity in their of the blood plasma was measured over here that is the CTQ over here. So, one is dark chocolate, one is dark chocolate in combination of milk chocolates, and one is completely milk. So, 12 subjects were used; 7 male and 5 females within the age groups and weights and body mass index like that.

On different days, a subject consumes one of the combination, and one hour later the total antioxidant capacity of that is measured. So, this is the observations that we are getting over here. We want to just check that whether there is any significant difference between the mean antioxidant level over here.

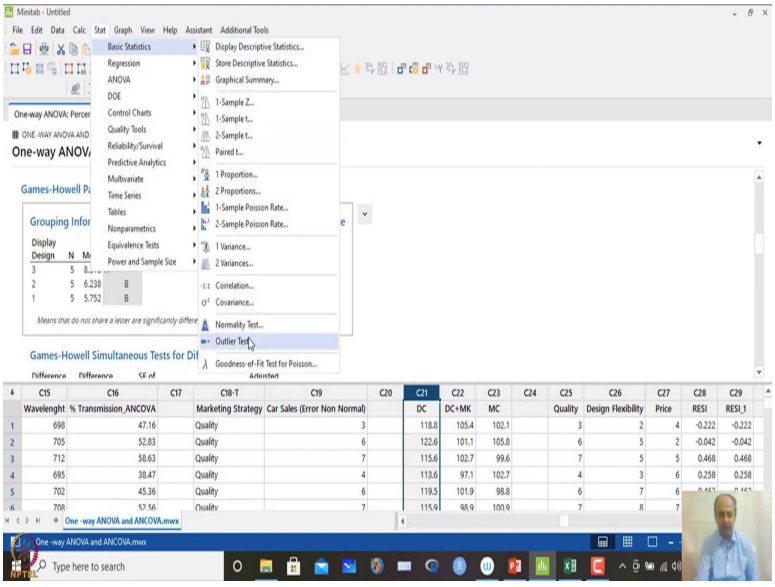
So, whether antioxidant capacity increases, if I only use dark chocolates like that, so that can be checked. And for this, these are given in columns information. So, when you do MINITAB ANOVA analysis, you test for equal variance.

(Refer Slide Time: 20:34)

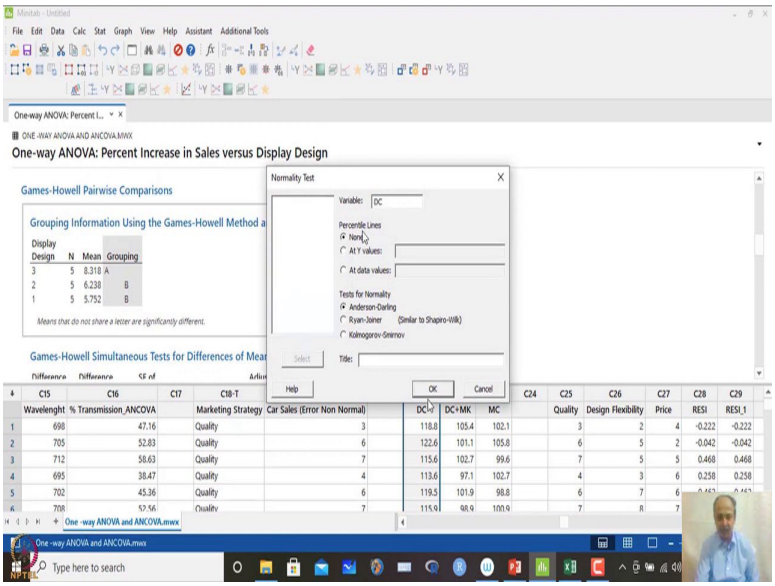


So, in this case what happens is that they are not in one column; you have to mention that they are in separate column. So, then you mentioned that which are the three different columns we are mentioning over here. Let us try to see whether they follow normal distribution or not.

(Refer Slide Time: 20:52)

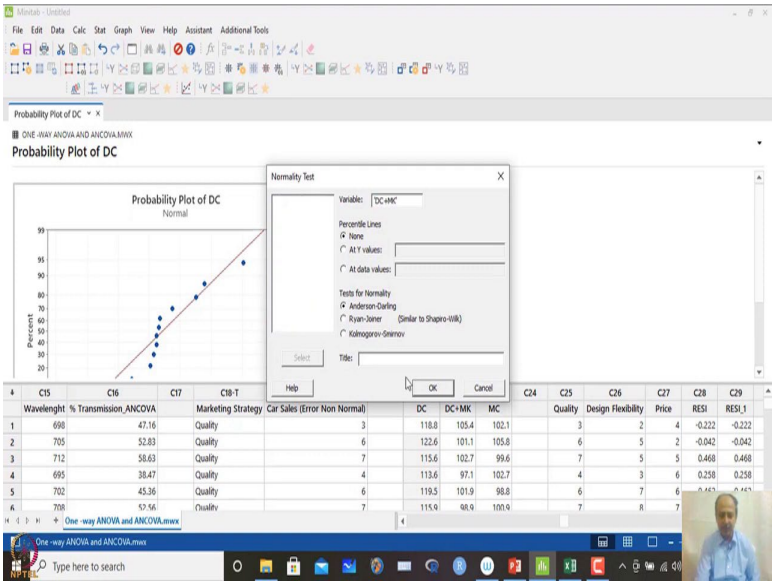


(Refer Slide Time: 20:54)

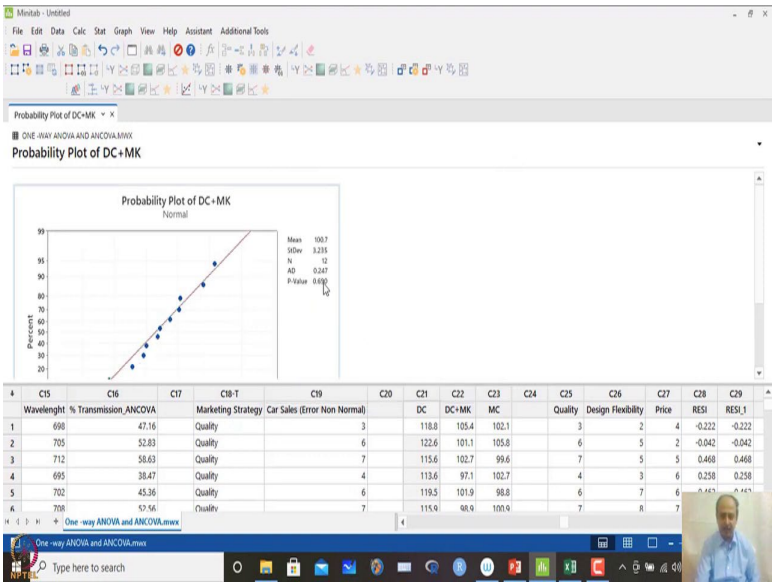


So, that we are confirming this one. So, in this case, we will go by dark chocolate and try to see whether it is normal.

(Refer Slide Time: 20:58)

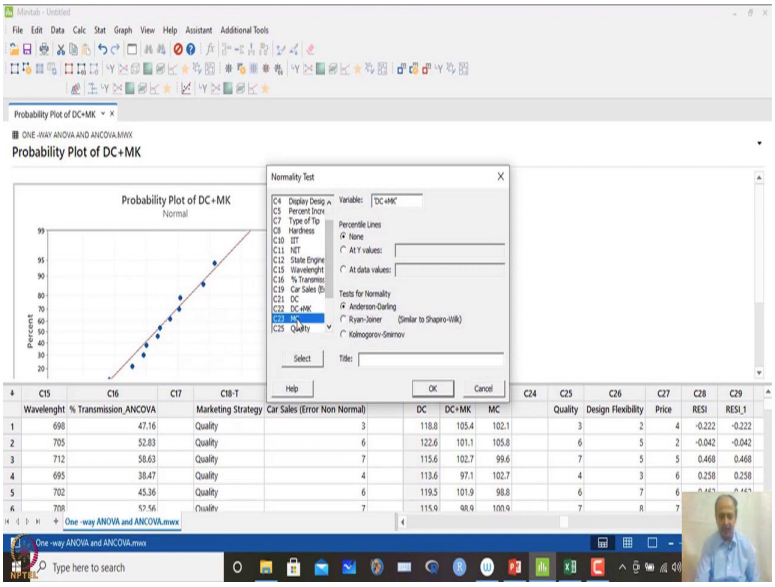


(Refer Slide Time: 21:12)

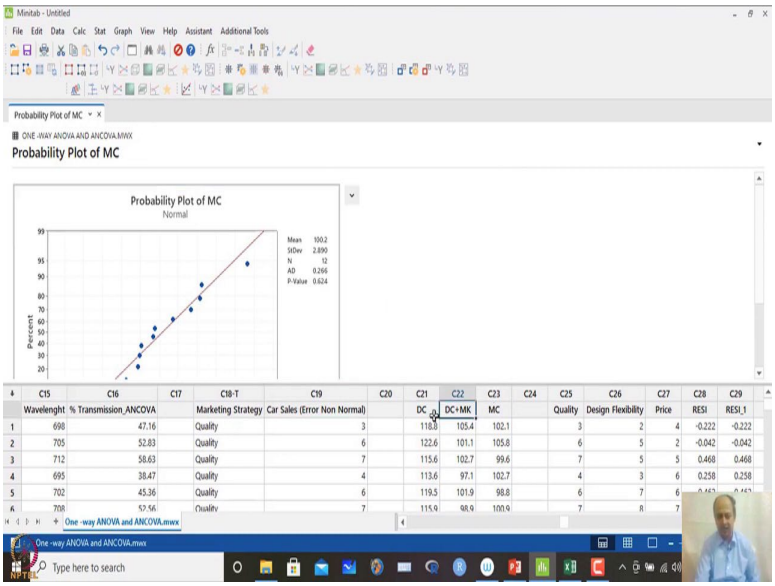


So, 0.07 is the and our level is 0.05. And we will we safely assume that this is normal. Second case, again I go to the next one. So, DC plus milk over here, and let me do the Anderson-Darling test. So, this is 0.69 this is also the group is having normal distribution assumptions is true.

(Refer Slide Time: 21:22)

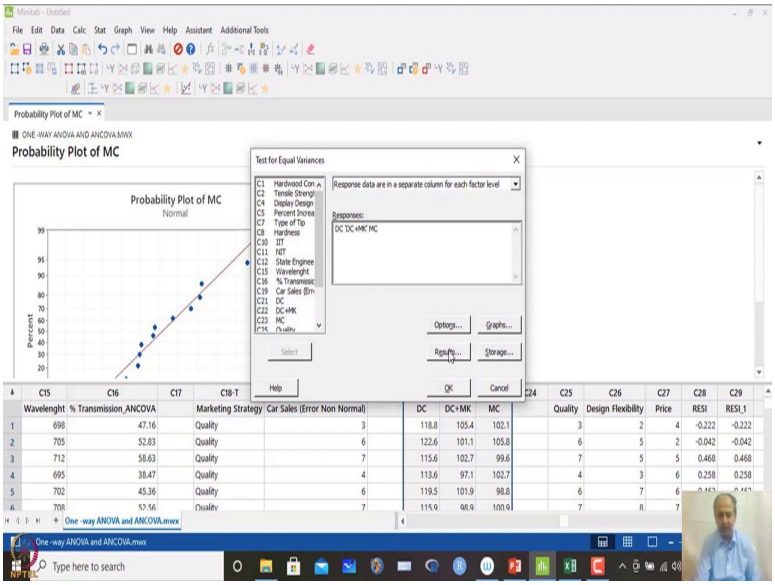


(Refer Slide Time: 21:27)

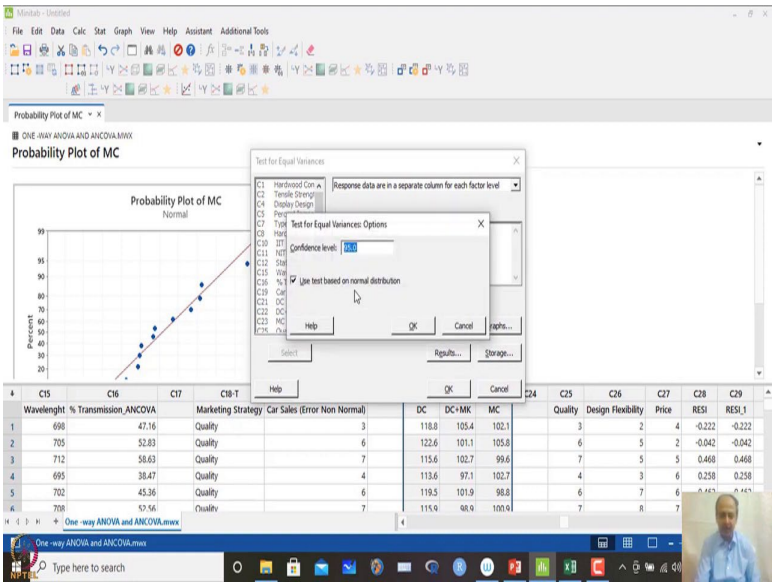


And the third one is the DC, only milk chocolate like that. Let me check with the assumptions. Here also it is not violating the normality assumptions over here.

(Refer Slide Time: 21:32)

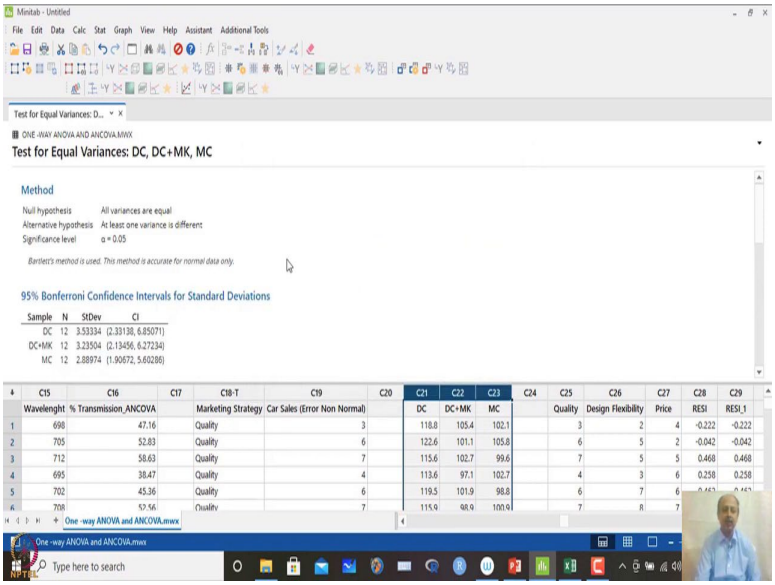


(Refer Slide Time: 21:44)

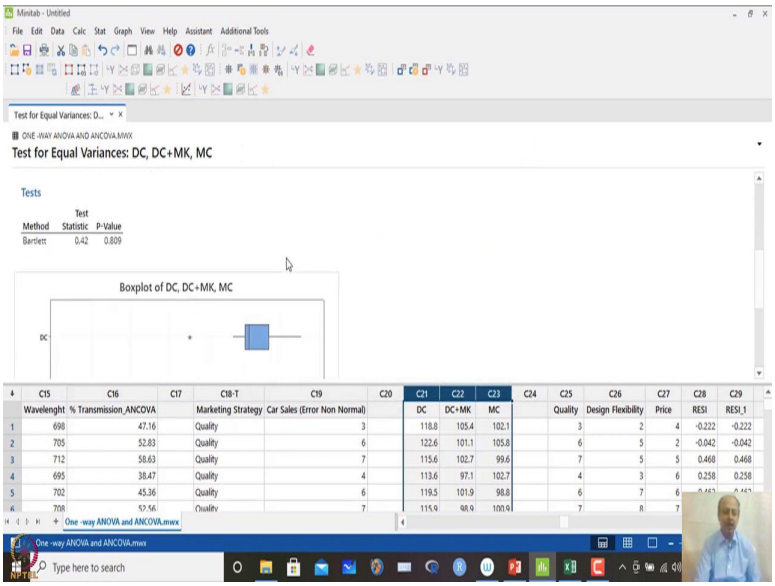


So, all the three columns are normality assumptions are fulfilled over here. And in this case what we can do is that *ANOVA*, *equal variance test*. So, in this case, they are in different column. So, we will mention that we want to see DC, DC+MK and this one is MC. And our options we will write use test for normality test, so Bartlett test will be used to confirm this one.

(Refer Slide Time: 21:48)

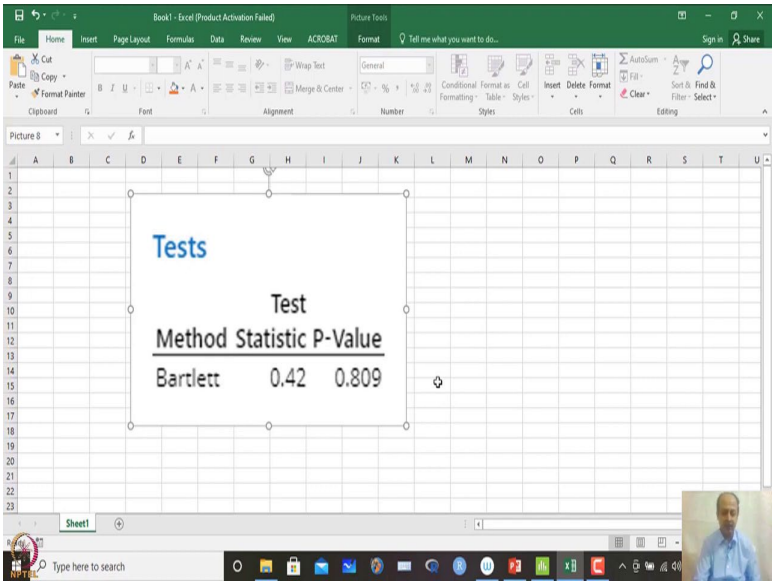


(Refer Slide Time: 21:50)



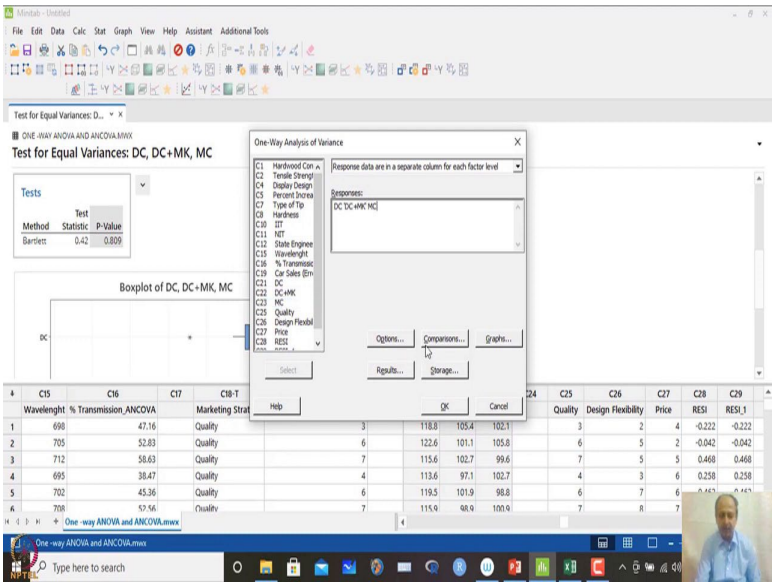
If you click ok, what will happen is that you will get the Bartlett test information over here and if I copy as picture, and I can paste it over here.

(Refer Slide Time: 22:00)

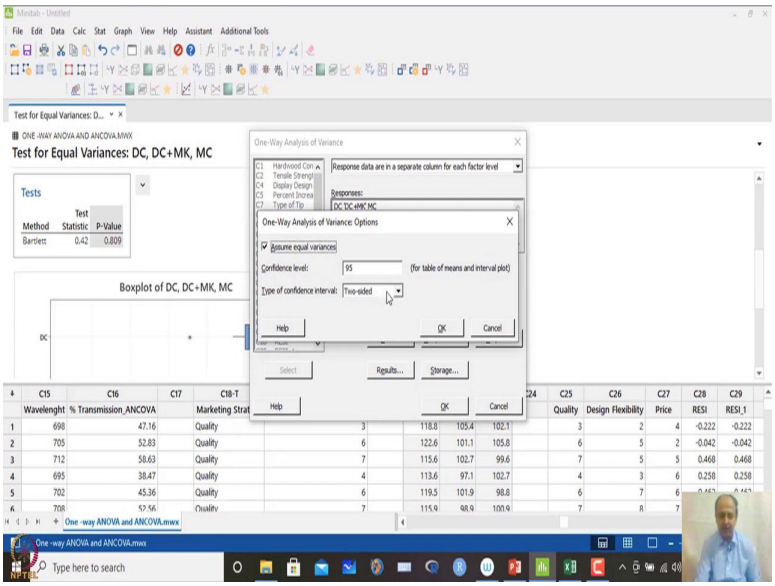


And when I paste this one information, what I see is that Bartlett test says that 0.08. And in this case confirms that there is no difference between the or variance are same at all levels variance are same over here. So, in this case, variance is same. Now, I have to do the ANOVA analysis and confirm.

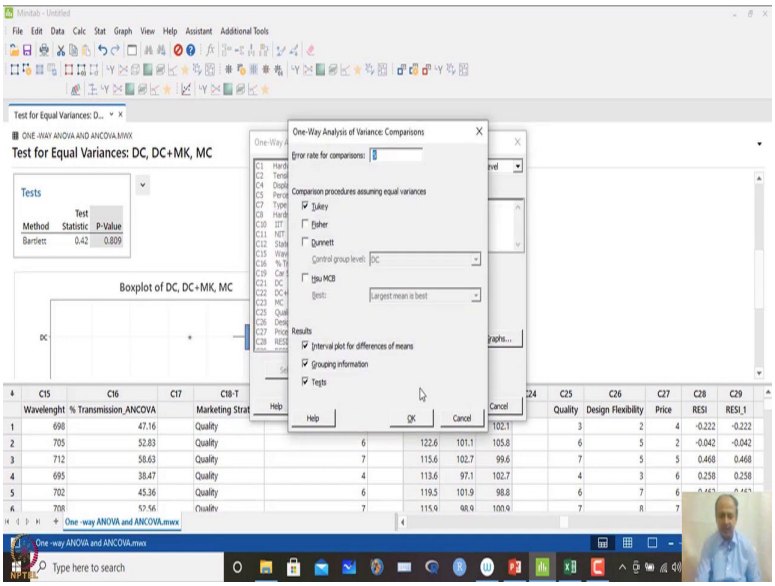
(Refer Slide Time: 22:24)



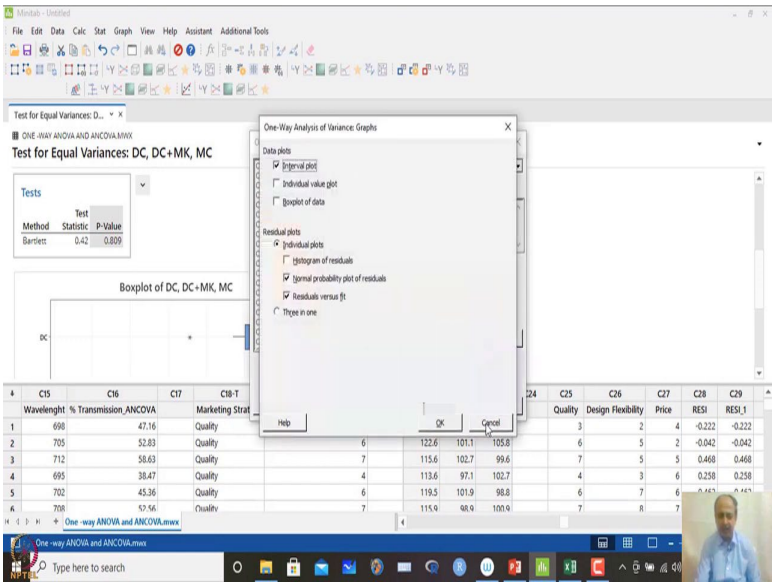
(Refer Slide Time: 22:37)



(Refer Slide Time: 22:41)

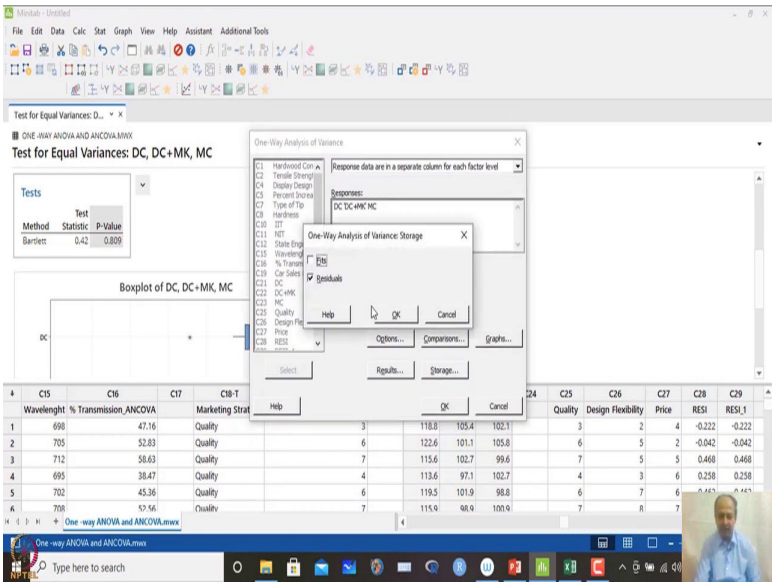


(Refer Slide Time: 22:46)



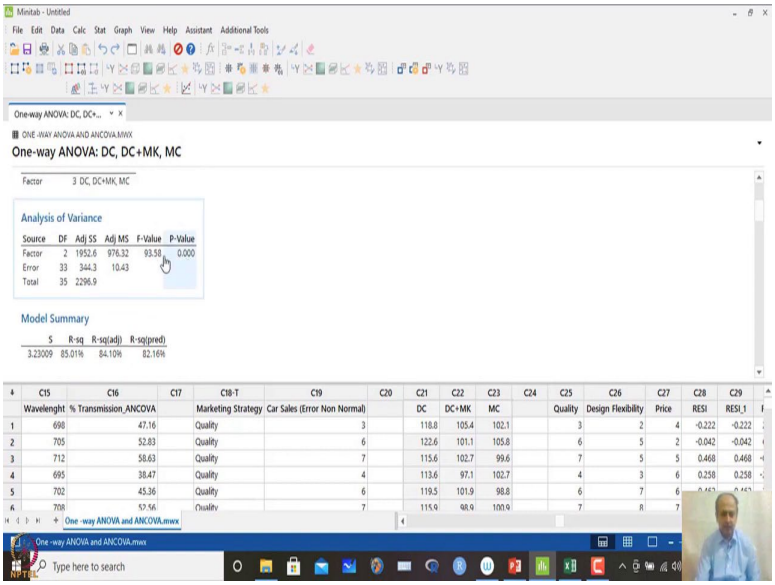
So, *ANOVA* analysis, *one way ANOVA* analysis, so everything is same. I will mention that they are in different columns. In options, we assume equal variance then Tukey's comparison test can be used. And the same thing can be seen graphically and we can save the residuals.

(Refer Slide Time: 22:48)

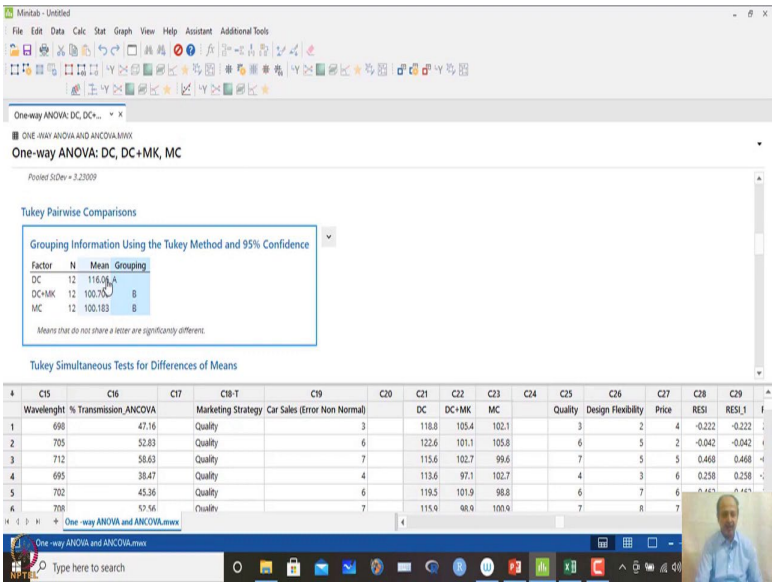


So, if we store the residual, then we can do the residual. So, graphically, we can see what is normal probability plot residual versus fit.

(Refer Slide Time: 23:06)



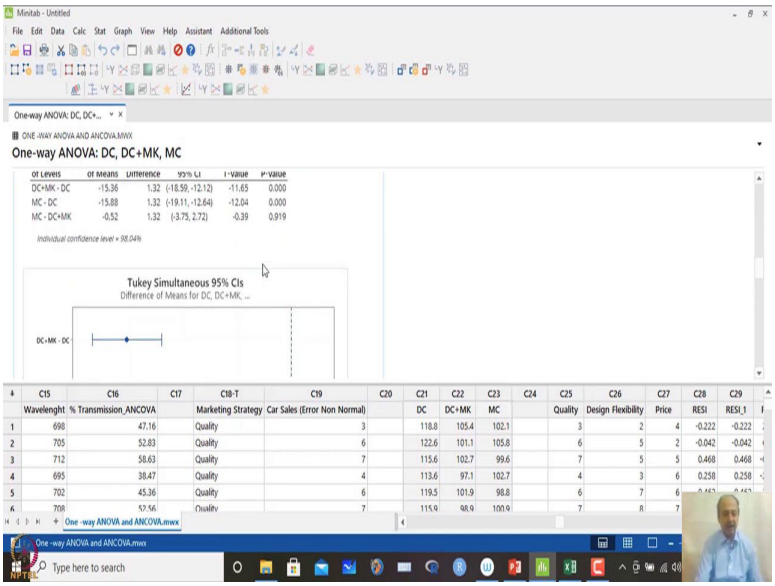
(Refer Slide Time: 23:21)



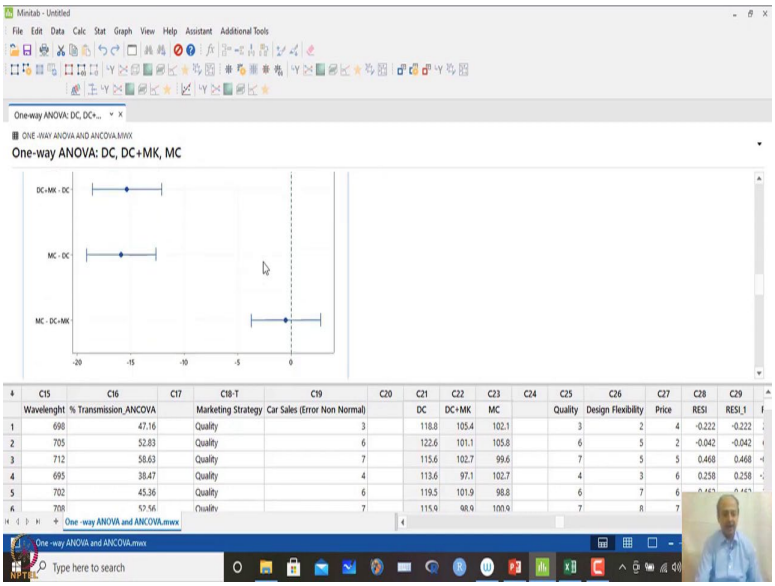
So, when we are doing this, what happens is that, they have significant difference which is confirmed in ANOVA analysis, p-value is close to 0. And which is different from which one so that is given in this letter code.

Dark chocolate is giving me higher mean anti antioxidant level that we are looking for antioxidant capacity of the blood plasma which is higher the better type of functions. So, if it is more, it is better. DC is having a group that is very different from the other groups. So, only with dark chocolate this seems to be giving me a higher mean which is what we are expecting.

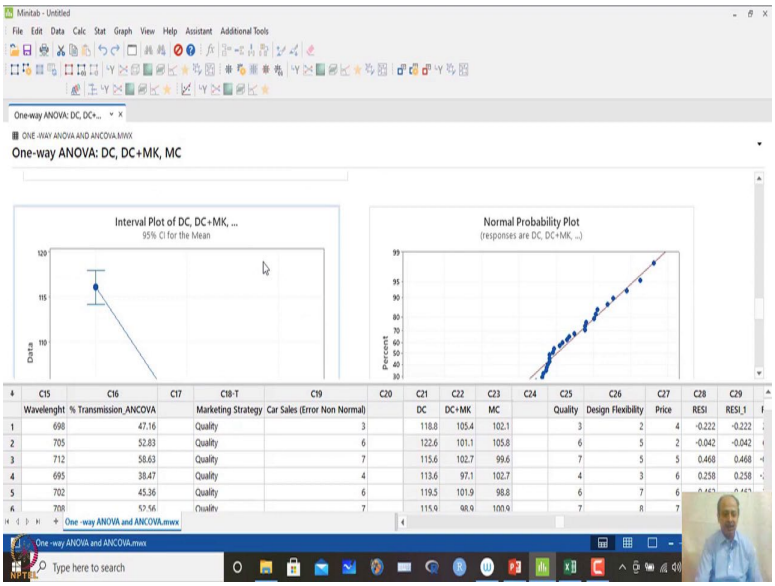
(Refer Slide Time: 23:47)



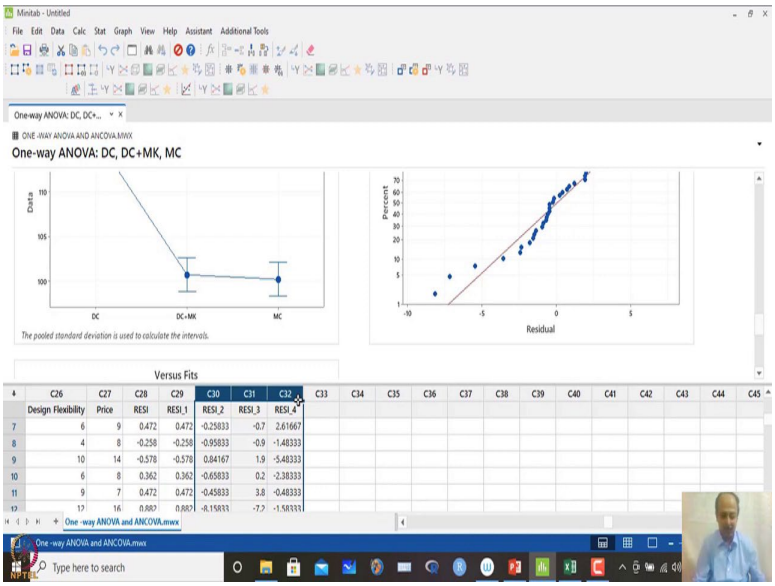
(Refer Slide Time: 23:47)



(Refer Slide Time: 23:48)

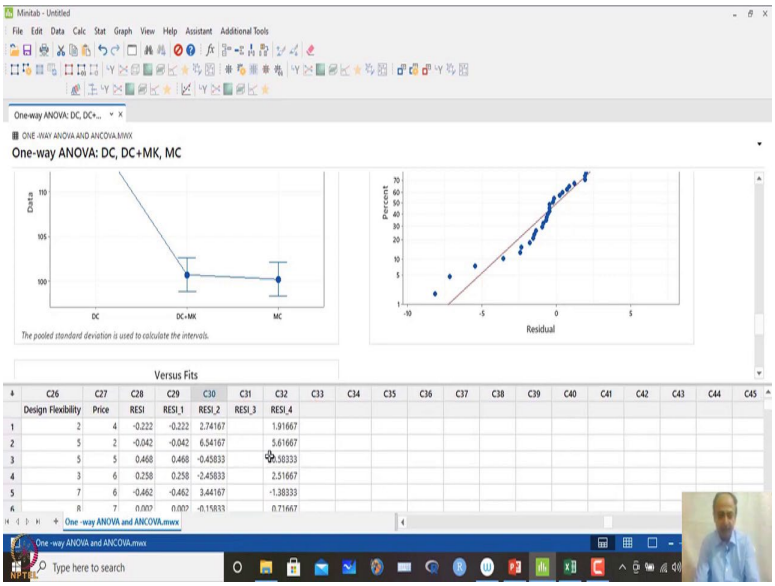


(Refer Slide Time: 23:48)



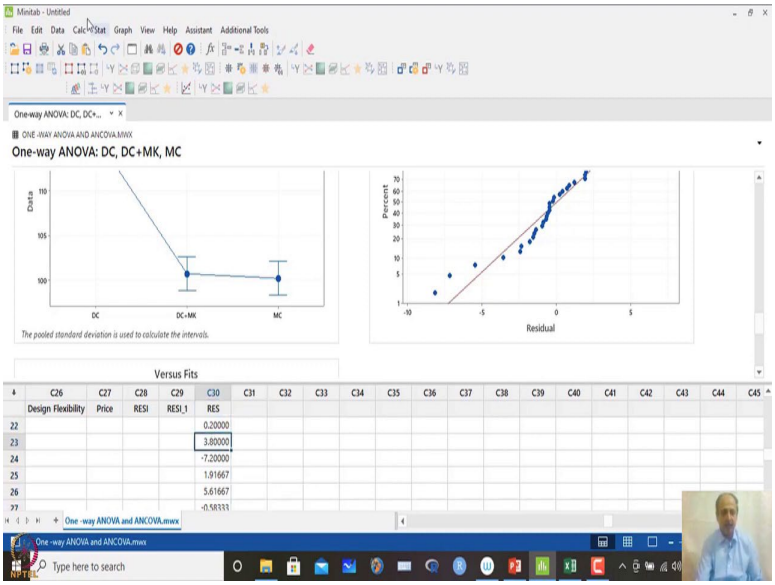
And then Tukey's test, then we have a normality test normally plot over here, does not seems to be violating that one condition. And the last column that you see over these are the three columns of residual over here. And we can check the normality assumptions over here. Each of the residuals we can check or we can combine all the residuals and do the checks and see whether the residual check is whether there is any violation in the assumptions like that.

(Refer Slide Time: 24:17)



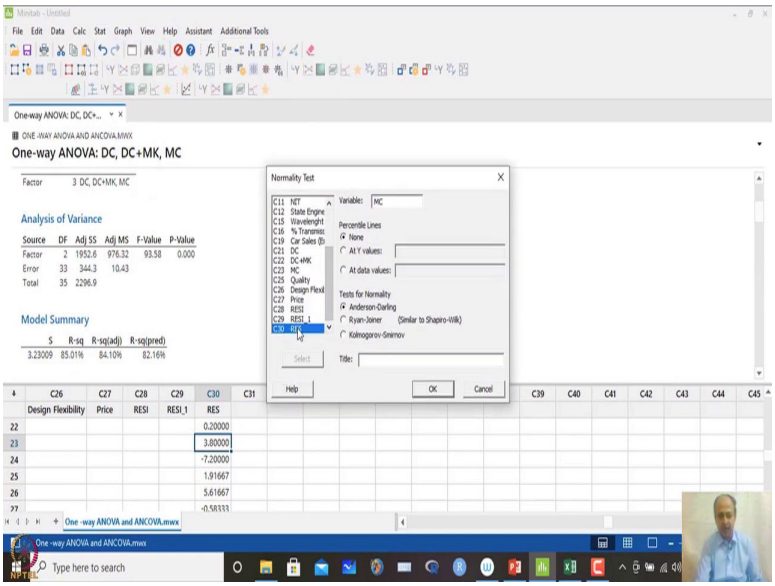
So, what we can do is that we can combine this and copy this one, control+x and we can paste this over here.

(Refer Slide Time: 24:29)

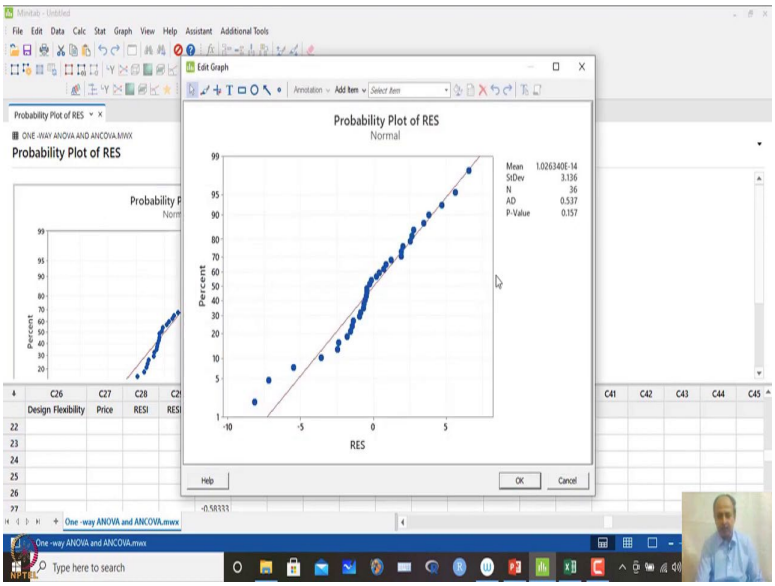


And we can just paste it so residual should be combined and in this case we can paste it over here. And we can remove these two columns and say this is the residual we are looking for. And we want to analyze this residual over here.

(Refer Slide Time: 24:39)



(Refer Slide Time: 24:44)

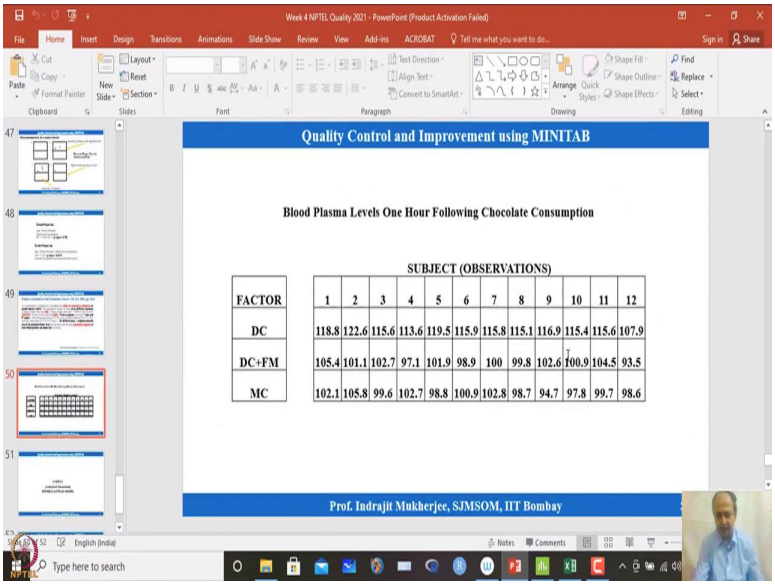


So, in this case, what we can do is that we can just check *statistic*, *basic stat*, *normality test* can be done over here, and the last residual column will be taken. Anderson-Darling test will be done. p value is 0.157. Hence there is no as such deviations of the normality assumptions like that.

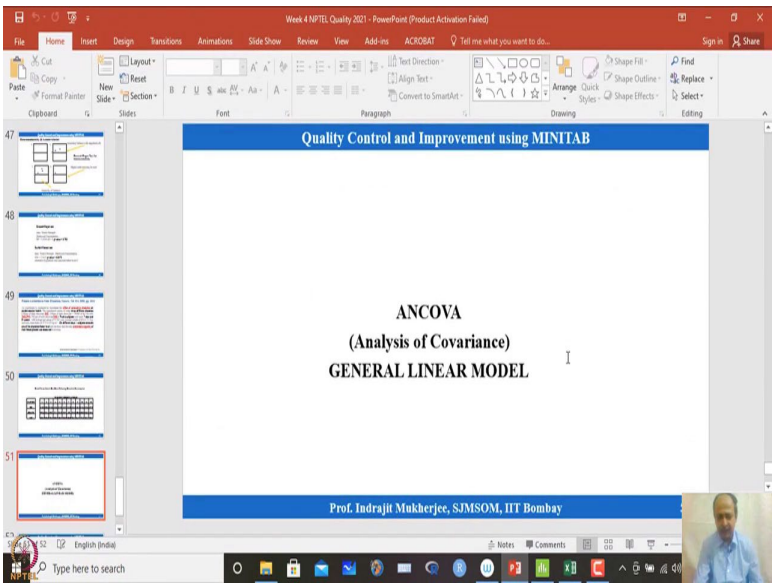
The other two checks Breusch- Pagan test can also be done in R and also we can do Durbin-Watson test statistics and compared with the table or we can see the p values

reported in R. And immediately, we can say whether autocorrelation or independency of the errors are true or not, so that can be verified.

(Refer Slide Time: 25:13)

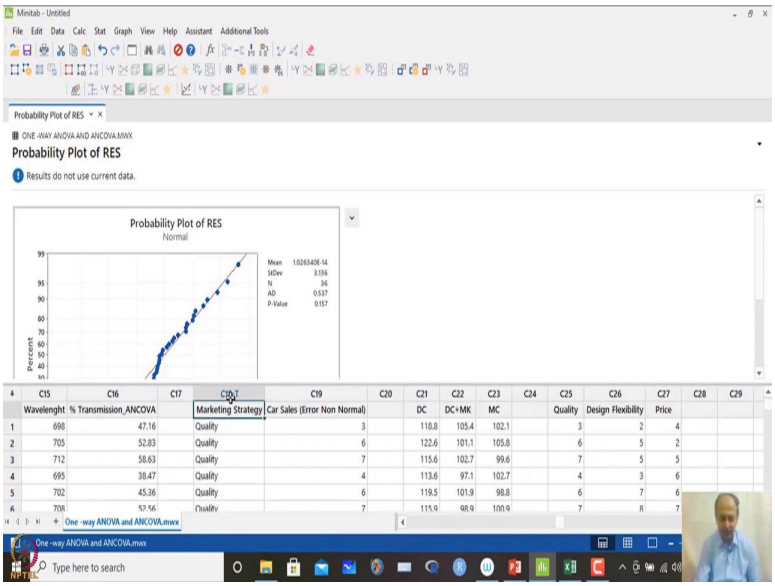


(Refer Slide Time: 25:14)



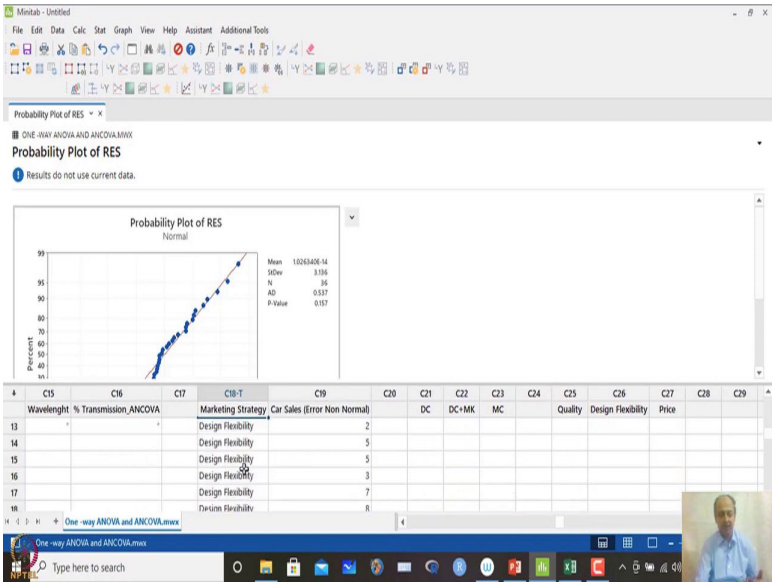
So, that are the conditions, but there can be scenarios there can be scenarios when we have assumption fails, when assumptions fails like that. So, let us take one example, where assumption fails. So, in this case, let me just show you where the assumptions may fail. And in this case, we may have to do the correction over here.

(Refer Slide Time: 25:34)

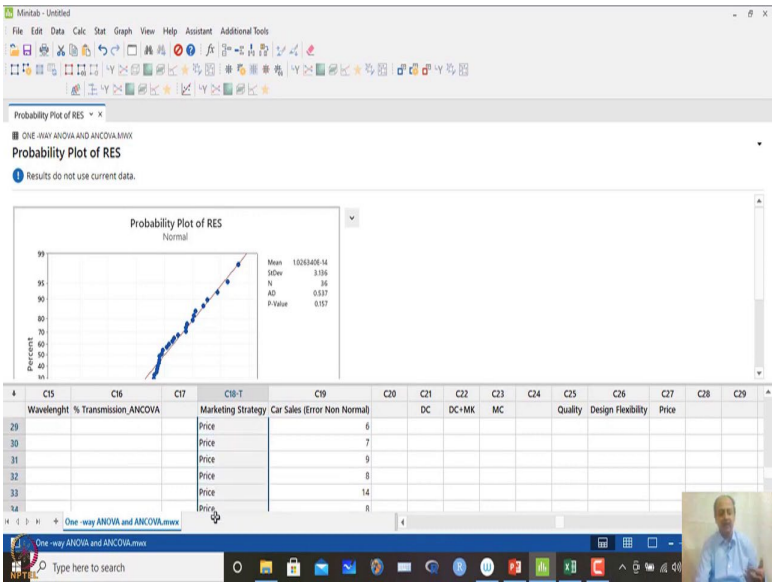


So, this is another example which I am using over here is that different marketing strategies taken and car sales are reported. And number of car sales let us assume that is continuity over here. So, in this case, although this is not continuous variable, but let us assume this variable is continuous over here.

(Refer Slide Time: 26:02)



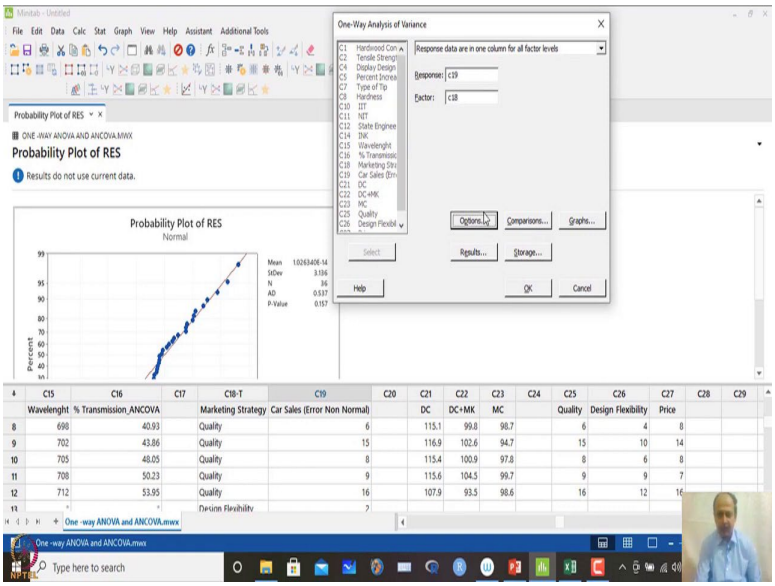
(Refer Slide Time: 26:08)



And we are going ahead with analysis of assumptions of ANOVA. And there are different way of advertisement that they have used. Sometimes they have given the agenda of quality as the advertisement agenda, sometimes they have talked about design flexibility in the advertisement, and sometimes they have given price as the priority while making the advertisement.

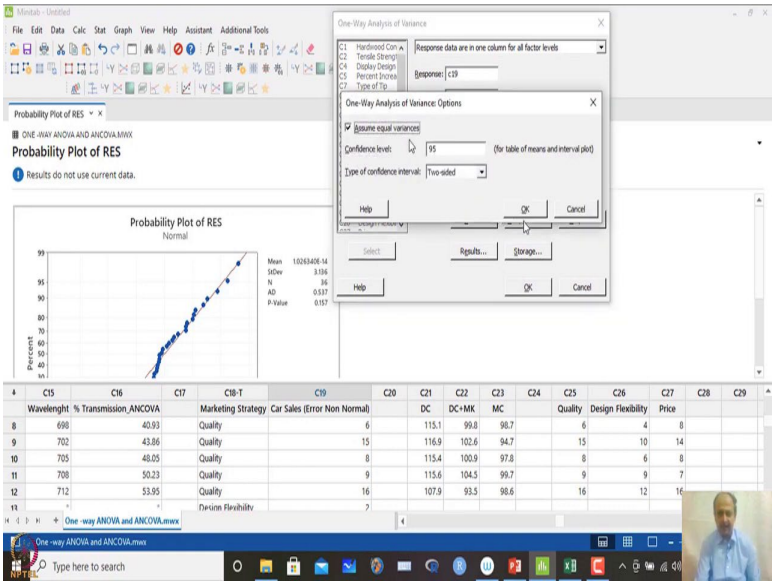
So, number of sales that is reported based on different types of advertisement at different periods are noted down. And then we want to see which is giving me higher car sales. So, in this case, ANOVA analysis can be used straightforward. And when I do this ANOVA analysis over here, let us assume all other conditions remains true.

(Refer Slide Time: 26:32)



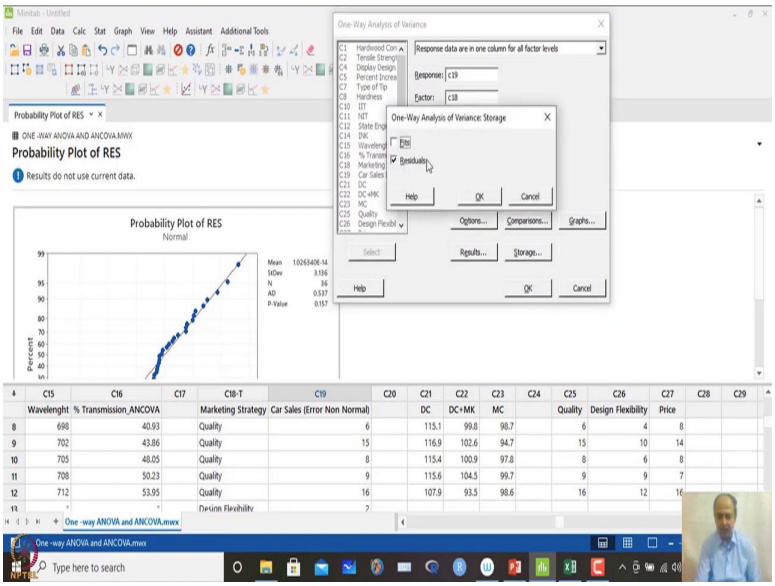
So, this is column number over here is C18 and C19 . Response is in C19 let us say, and strategy is in C18.

(Refer Slide Time: 26:57)

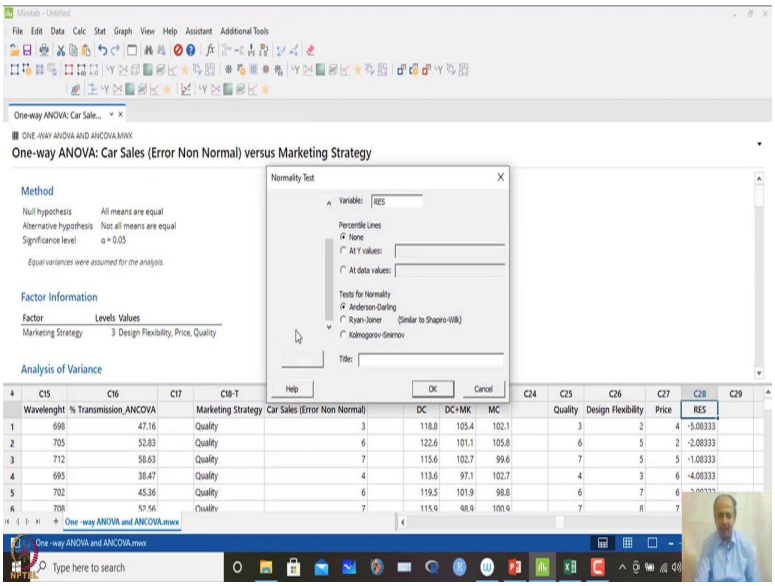


And options is that assume equal variance this we can check, so maybe afterwards we are checking like that.

(Refer Slide Time: 27:05)

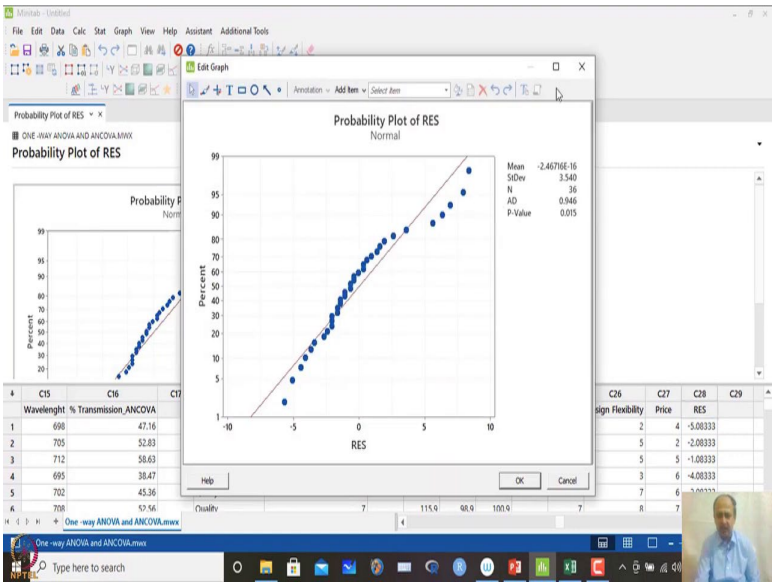


(Refer Slide Time: 27:09)



So, here what I am interested in to show that the residual we have saved at the end, and let us run the ANOVA analysis. So, this is the residual last residual that you see will be the residual of the ANOVA models that we are getting. And when this is the residual, let me just cross check whether this is normally distributed or not.

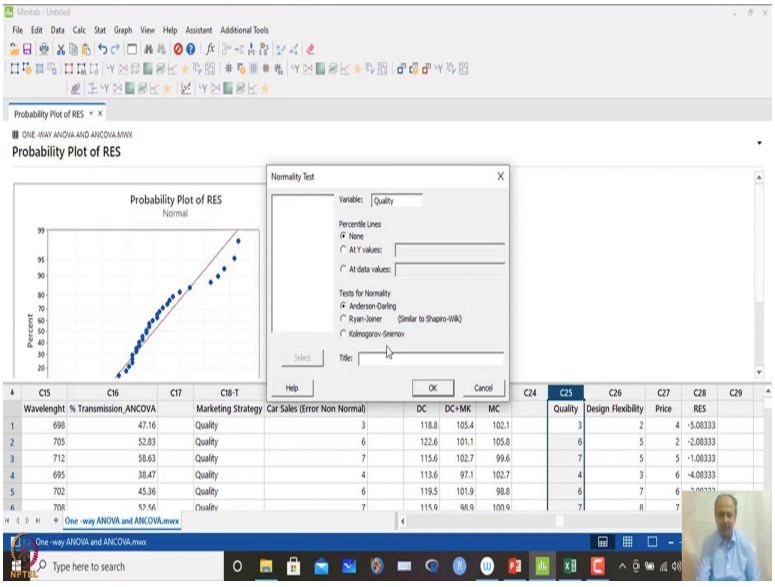
(Refer Slide Time: 27:27)



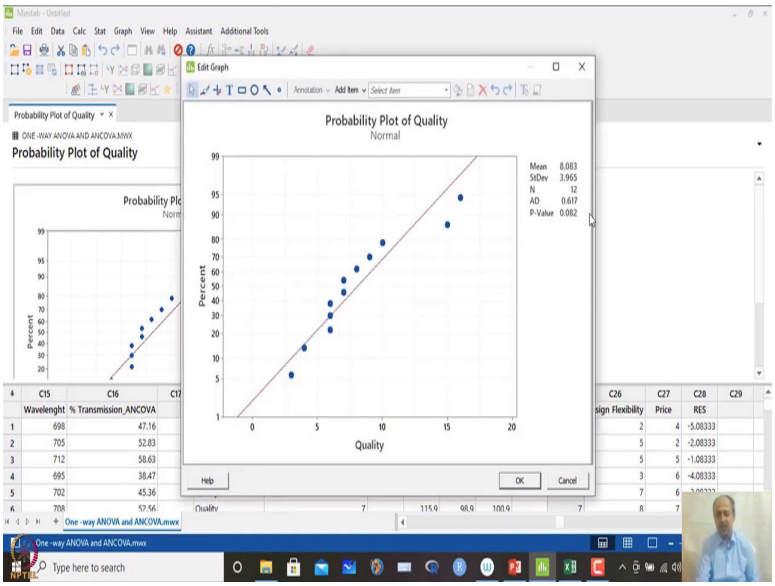
So, when you go for the residuals at the end and click this one, and do Anderson-Darling test what happens is that you observe that the residuals are non-normal, because p value is less than 0.05, here it is reported as 0.015. So, then in that case, what is to be done?

If the if some condition fails or whether it is heteroscedasticity, whether it is other condition independency assumptions over here and whether it is normal distribution assumptions that fails ok. So, when I have done individual testing over here, let us say this is quality design flexibility and price, and they are in different groups.

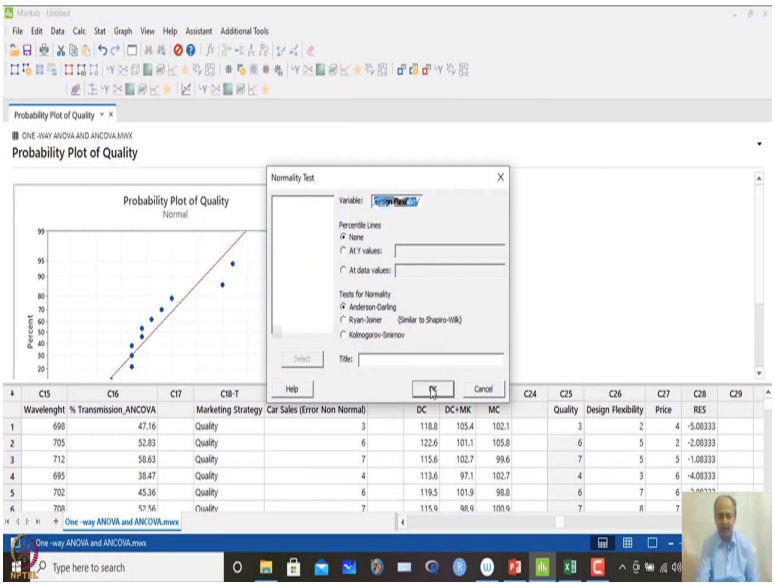
(Refer Slide Time: 28:03)



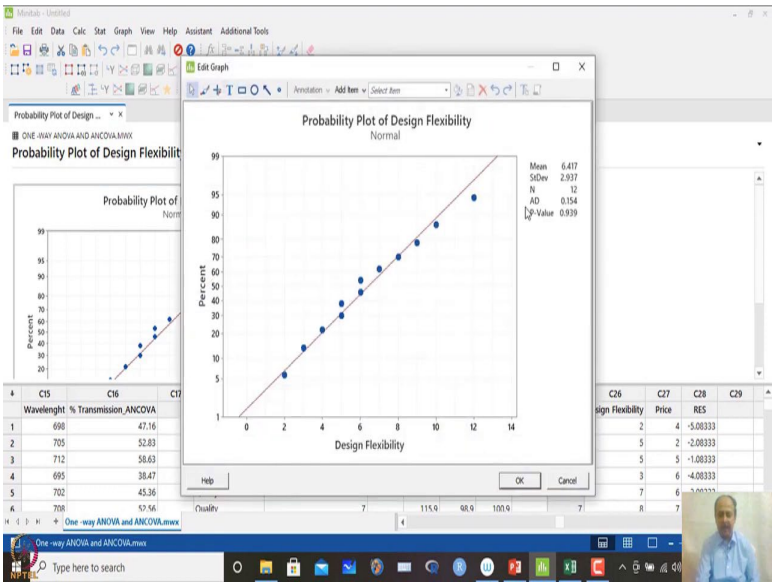
(Refer Slide Time: 28:09)



(Refer Slide Time: 28:18)



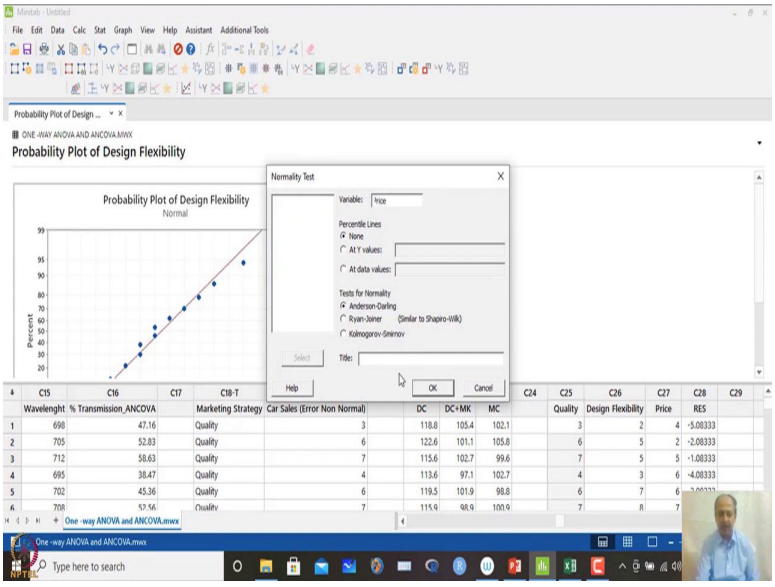
(Refer Slide Time: 28:19)



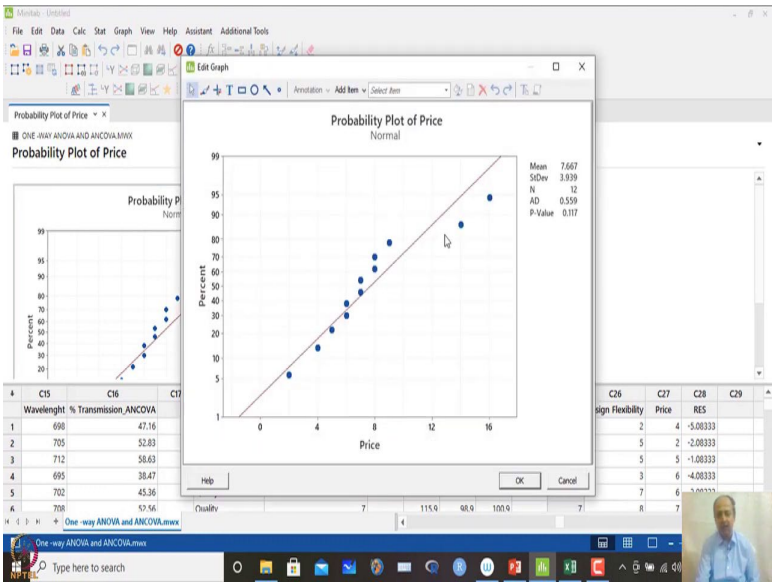
When I do only on the y whether they are normally distributed or not, see if I have done the individual observations I have taken and I have done this test let us say for quality, what I see Anderson-Darling link test is more than 0.05.

So, this is satisfactory and then if I go to the second observations and I do the normality test design flexibility and do the Anderson-Darling test, again I see 0.939, this is also satisfactory. And the third level when I go ANOVA analysis, sorry, this basic statistics and normality test over here.

(Refer Slide Time: 28:33)

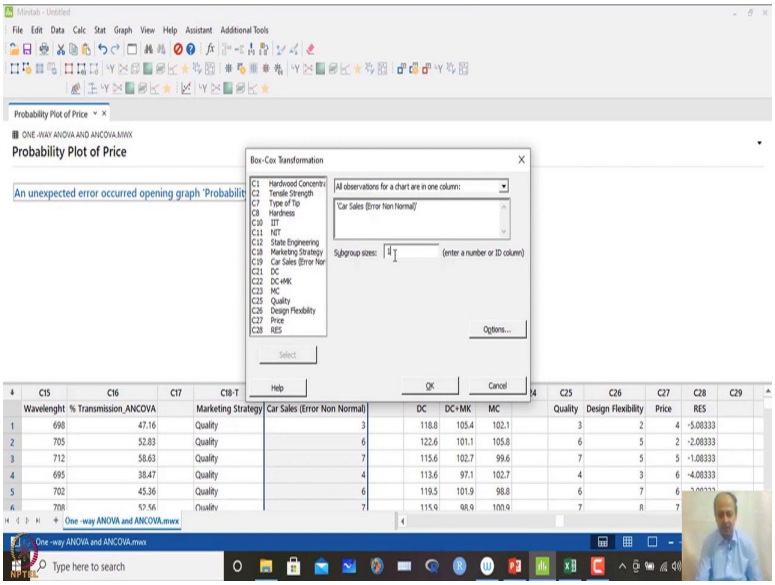


(Refer Slide Time: 28:35)



And do the last and final one groups that is price. And I want to see whether the y variable is normal or not in that group. So, this is 0.117, this is also normal.

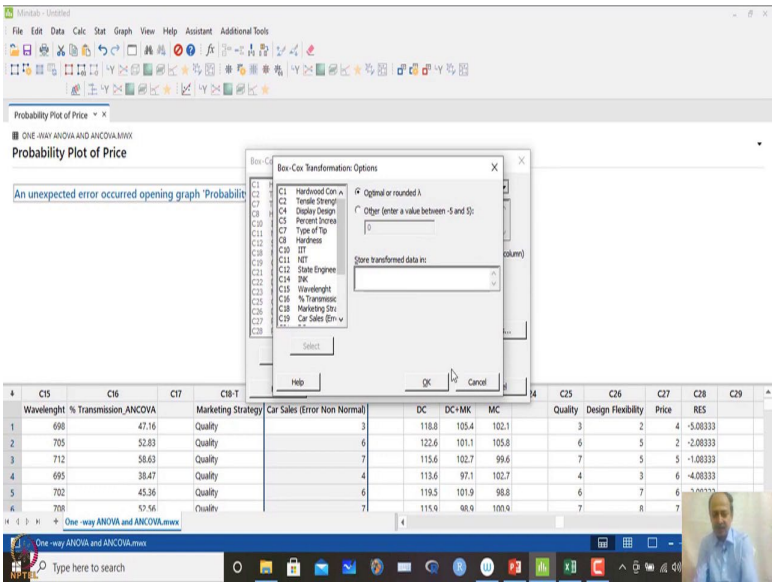
(Refer Slide Time: 28:40)



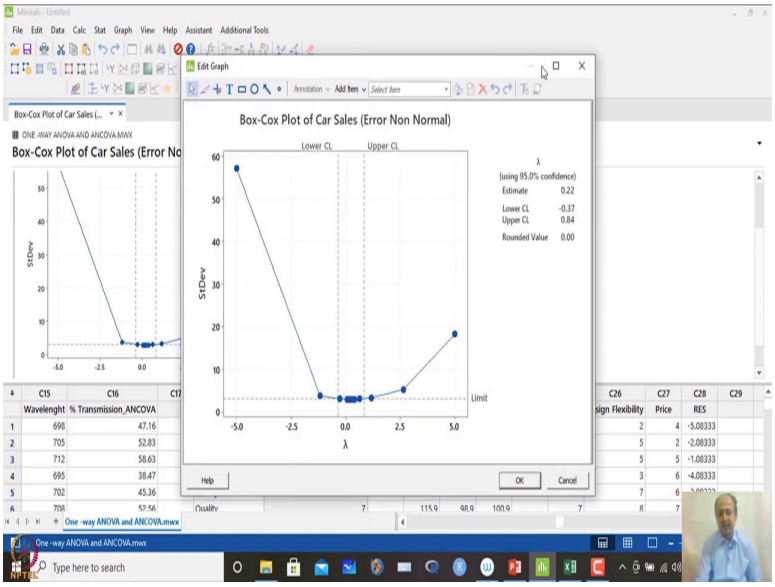
So, you see y is normal over here. But when I do the ANOVA analysis and save the residual, residual is not following normal distributions like that. So, we have to always check the residual and confirm that whether the normality assumption or the assumptions of the ANOVA analysis depends on all on residual basically.

So, I have to check the residual and see the conditions. And if this is not true what can be done is that we have to transform the y variables, which is the CTQ like that. One of one thing what we can do we have discussed already is that we can convert this into using Box-Cox transformation over here.

(Refer Slide Time: 29:20)



(Refer Slide Time: 29:31)

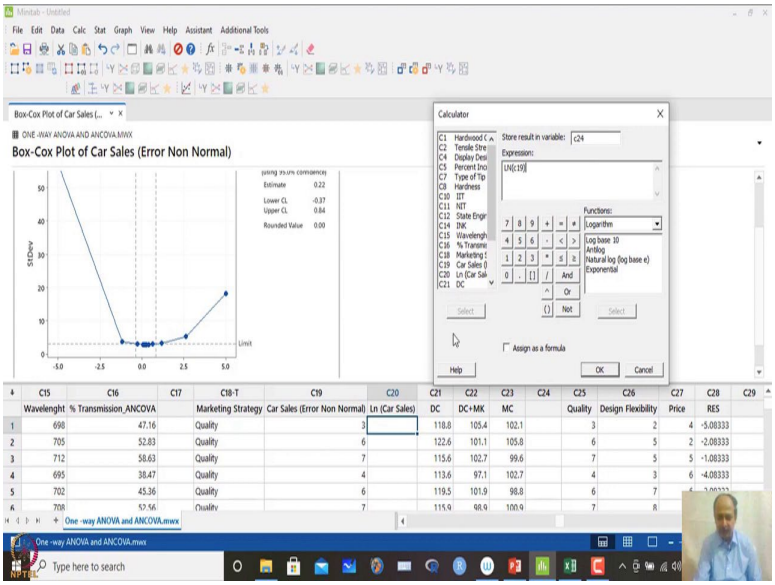


So, I want to change this variable and make a transformation over here. I have to select a subgroup size of one. And in options what I have given is that I will go by so, I am not saving this information at present. Let me see what is the optimal Box-Cox

transformation that it gives and it is saying that use a transformation of rounded value of 0.

So, lambda transformation of 0 indicates ln transformation that is natural logarithm transformation over here. So, ln transformation base e will be used over here. And in this case what we can do is that we can make it over here.

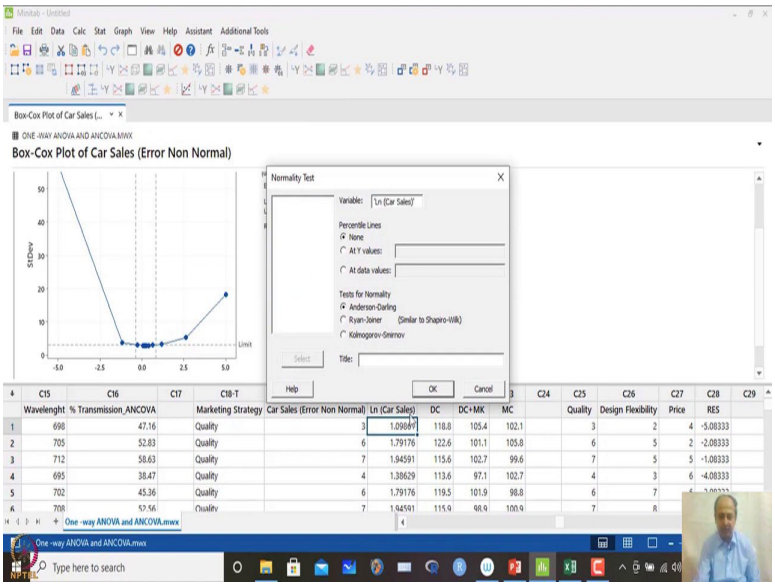
(Refer Slide Time: 29:50)



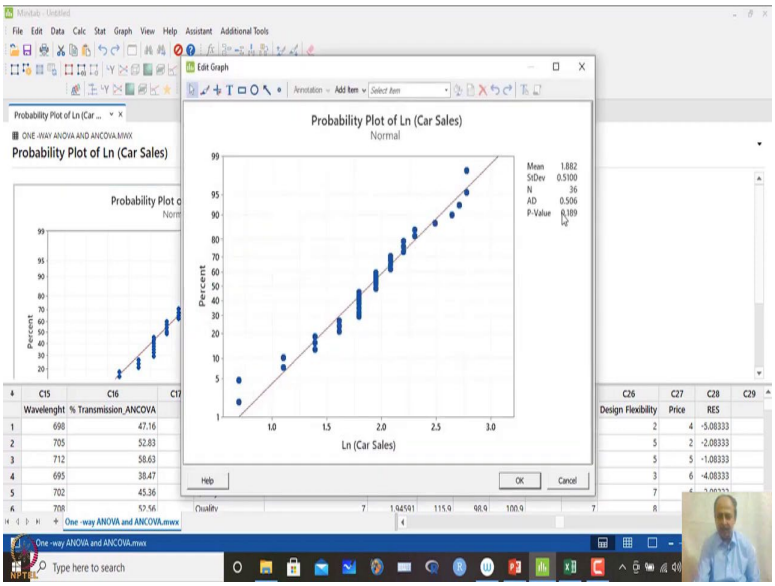
So, we can write that ln of car sales let us say and we can just calculate the values over here. So, we can use calculators over here. And we can just choose ln information. So, ln of C19 I have put over here, so ln information, I store it in C20.

So, expression is C19 we are using the conversion. And if I click ok this is the transformation that is done on this data set that we are having. Let us check whether transformation has changed into normal or not y into normal assumptions whether it is fulfilled like that.

(Refer Slide Time: 30:31)

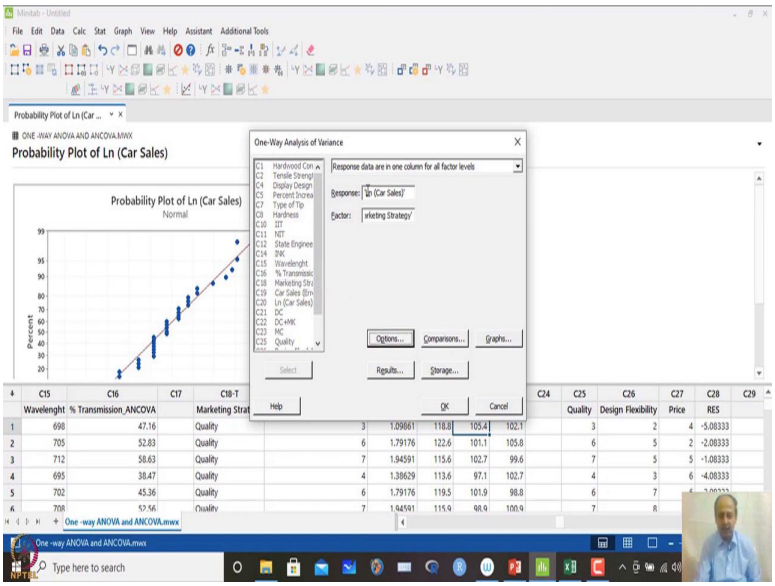


(Refer Slide Time: 30:33)



So, ln car let me just try to check and it is showing that this is following normal distribution. So, there is no problem. So, let me do the ANOVA analysis over here.

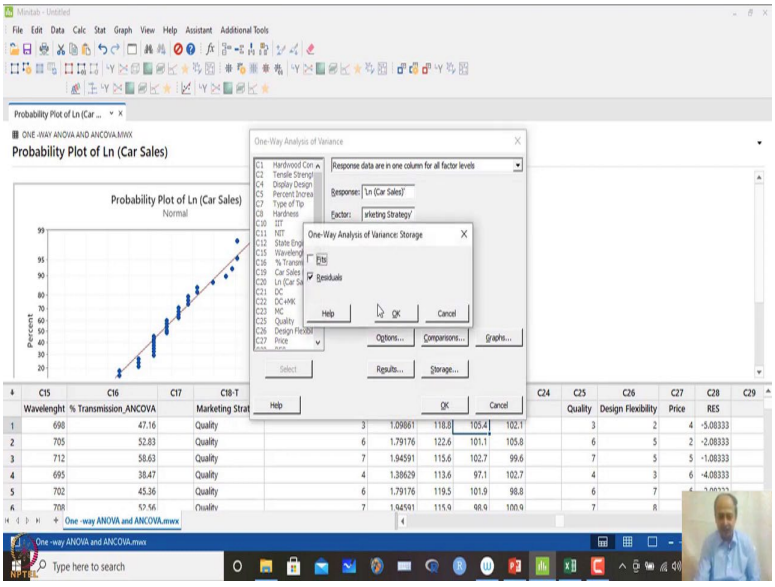
(Refer Slide Time: 30:42)



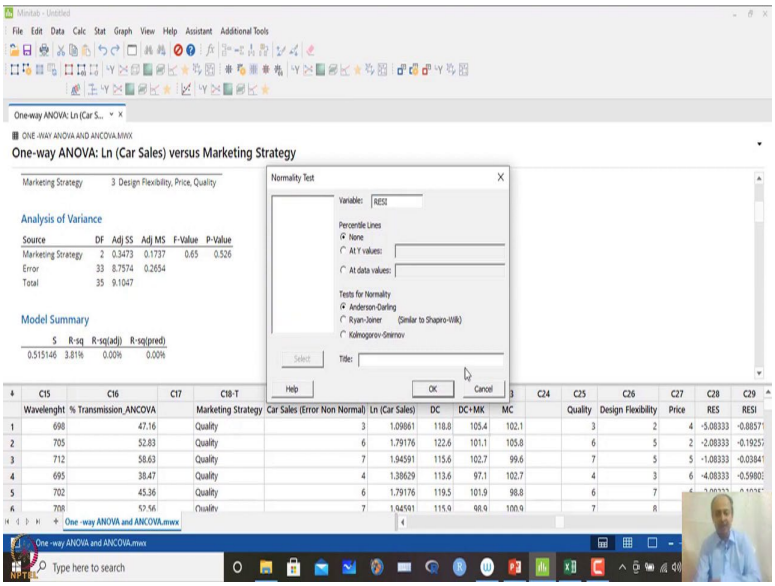
So, in this case, what we have to do is that response we have to change. So, I have to change because I have transformed the data, because of that non-normal residuals that we are getting. So, I have used a conversion over here which will be used.

And on the converted y, I will do the one way analysis of variance. So, in this case, the factor will be same only the y condition, y will be changed to y lambda which is lambda is over here is 0, and that is that indicates a logarithmic transformation over here, natural logarithmic transformation.

(Refer Slide Time: 31:18)

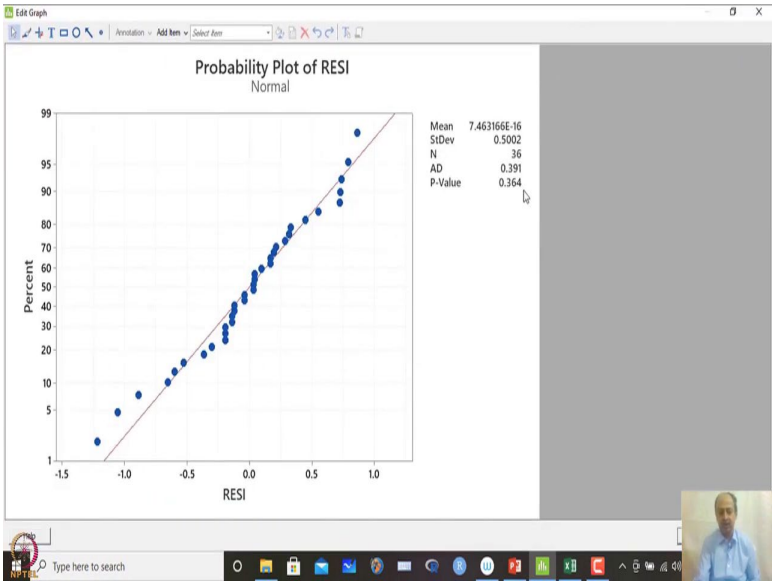


(Refer Slide Time: 31:23)



And then I will again store the residual at the end, and I will try to see what happens. So, here also it is saying that there is no difference in marketing strategy, because p value is not significant. So, marketing test strategy does not influence basically the outcomes or car sales like that. Now, let me check the final residual whether it is now normal or not. So, I will do the normality test on the final residual which is residual one which is saved at the final end.

(Refer Slide Time: 31:43)



So, if I click ok over here, what will happen is that you see now the residual is following normal distributions like that. So, whenever I have done the transformation on y, what happened is that and then I have run on the transform y data and the x factors that is same ok, then I will get the true information.

So, sometimes what happens is that if you ignore the model adequacy test, in that case what may happen is that you may be concluding wrongly in certain scenarios like that. Misleading information may come out of the ANOVA analysis.

So, proper transformation may be used over here, but although ANOVA is quite robust to handle any non normal behaviors of the residual. Although people says that or statistician says that this technique is quite robust. And even if some deviation model deviation in the normality assumptions or model adequacy checks are not correct, in that case also whatever results you get may be quite adequate to make a conclusions based on that ok.

So, all these checks needs to be done, because model adequacy check is an important aspect. And if you can correct that one and do the ANOVA analysis that, then you can be more sure that whatever conclusions that you are drawing of the ANOVA analysis or comparison test like that is quite correct, and can be generalized like that for the given levels that are there.

So, this one way ANOVA analysis is basically when we are trying to see the transformation, when we are just shifting from control phase to improvement phase, and trying to determine which are the which are the factors needs to be considered in experimentation.

So, this type of small small one factor analysis of one way one factor one way analysis of variance can be used, and then T test, two sample T test, paired T test or this kind of small experimentation. And one important tool that can also be of help while selecting the factors over here is known as regression analysis that can be used to identify whether the x variable is when x is continuous and y is continuous, so x variable is also continuous y is also continuous.

So, those kind of variables can be best identified if you are if you are using regressions like that. So, simple previous historical data when experimentation was not done, some

previous data and some correlation some linear equations can be established. So, those things we can see when we talk about regression analysis which is extremely useful not only in design of experiment, but while segregating the factors also some of the previous information can be used.

And that will give you a lead whether the factors to be included or to be excluded from the analysis. So, we will discuss about regression in our next lecture, from simple regression to multiple regressions like that which will be helpful in our design of experiments.

And moreover, we will add over here analysis of covariance another one important aspects which we have missed out over here. So, we will start with analysis of covariance and then we will shift to regression analysis where x is also continuous and y is continuous, how to identify that whether x influences y or not.

So, those kind of things which will be used as a screening in the screening phase and which will be used in basically full flow experimentation in the improvement stage. So, these are the techniques which can be used. And MINITAB gives you all options to explore. So, we will stop here, and we will continue from here in our next session.

Thank you.